

Präzise Prognose von Störungen an Eisenbahnweichen

Von der Fakultät für Maschinenbau
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung der Würde

eines Doktor-Ingenieurs (Dr.-Ing.)

genehmigte Dissertation

von:

Dipl.-Ing.-Inf. Thomas Böhm

aus:

Magdeburg (Deutschland)

eingereicht am: 20. Dezember 2016

mündliche Prüfung am: 16. Mai 2017

Gutachter:

Professor Dr.-Ing. Karsten Lemmer

Professor Dr.-Ing. Dr. h.c. mult. Eckehard Schnieder

Berichte aus dem DLR-Institut für Verkehrssystemtechnik

Band 33

**Präzise Prognose von Störungen an
Eisenbahnweichen**

Thomas Böhm

Herausgeber:

Deutsches Zentrum für Luft- und Raumfahrt e. V.
Institut für Verkehrssystemtechnik
Lilienthalplatz 7, 38108 Braunschweig

ISSN: 1866-721X

DLR-TS 1.33

Braunschweig, im Februar 2018

Institutsdirektor:
Prof. Dr.-Ing. Karsten Lemmer

Verfasser:
Dipl.-Ing.-Inf. Thomas Böhm

Vorwort des Herausgebers

Liebe Leserinnen und Leser,

in Ihren Händen halten Sie einen Band unserer Buchreihe „Berichte aus dem DLR-Institut für Verkehrssystemtechnik“. In dieser Reihe veröffentlichen wir spannende, wissenschaftliche Themen aus dem Institut für Verkehrssystemtechnik des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR) und aus seinem Umfeld. Einen Teil der Auflage stellen wir Bibliotheken und Fachbibliotheken für ihren Buchbestand zur Verfügung. Herausragende wissenschaftliche Arbeiten und Dissertationen finden hier ebenso Platz wie Projektberichte und Beiträge zu Tagungen in unserem Hause von verschiedenen Referenten aus Wirtschaft, Wissenschaft und Politik.

Mit dieser Veröffentlichungsreihe verfolgen wir das Ziel, einen weiteren Zugang zu wissenschaftlichen Arbeiten und Ergebnissen zu ermöglichen. Wir nutzen die Reihe auch als praktische Nachwuchsförderung durch die Publikation der wissenschaftlichen Ergebnisse von Dissertationen unserer Mitarbeiter und auch externer Doktoranden. Veröffentlichungen sind wichtige Meilensteine auf dem akademischen Berufsweg. Mit der Reihe „Berichte aus dem DLR-Institut für Verkehrssystemtechnik“ erweitern wir das Spektrum der möglichen Publikationen um einen Baustein. Darüber hinaus verstehen wir die Kommunikation unserer Forschungsthemen als Beitrag zur nationalen und internationalen Forschungslandschaft auf den Gebieten Automotive, Bahnsysteme und Verkehrsmanagement.

Das Bahnsystem ist auf der einen Seite geprägt von einem wachsenden Wettbewerbsdruck durch andere Verkehrsträger und der steigenden Nachfrage nach zuverlässigeren, pünktlichen Verbindungen. Auf der anderen Seite nehmen die Instandhaltungskosten zu. Um diesen Herausforderungen zu begegnen, braucht es innovative Lösungen für die Instandhaltung.

Der vorliegende Band stellt einen datengetriebenen Ansatz zur präzisen Prognose von Störungen an Eisenbahnweichen vor, mit dem der Infrastrukturbetreiber mindestens 50 Prozent der Störungen und Verspätungsminuten verhindern kann. Der Bericht zeigt systematisch, welche Datenquellen dafür notwendig sind, achtet aber darauf, bestehende Weichen nachrüsten zu können und weitestgehend auf zusätzliche Sensorik zu verzichten. Auf der Basis umfangreicher Felddaten wird mittels Algorithmen der künstlichen Intelligenz die verbleibende Nutzungsdauer vorhergesagt. Erstmals werden dazu offene, externe Datenquellen verwendet und statische mit dynamischen Daten verknüpft, um die notwendige Genauigkeit zu erreichen. Die Frage nach der notwendigen Genauigkeit im unvermeidlichen Trade-off zwischen Fehlalarmen und unerkannten Störungen beantwortet die Arbeit aus wirtschaftlicher Sicht. Dies geschieht mit einer eigens entwickelten und vom Anwendungsfall unabhängigen Methode. Sie stellt sicher, dass nicht nur Störungen verhindert werden, sondern auch tatsächlich Kosten reduziert werden. Insgesamt leistet das beschriebene Verfahren einen wertvollen Beitrag für eine höhere Pünktlichkeit und geringere Kosten und stärkt damit die Schiene als umweltfreundlichen und sicheren Verkehrsträger.

Prof. Dr.-Ing. Karsten Lemmer

Vorwort des Autors

Bereits während meines Studiums faszinierte mich die Möglichkeit, Wissen aus großen Datenbeständen zu extrahieren. Ich bin glücklich, dass diese Faszination ein Teil meines Berufes ist, denn sie war die stärkste Triebkraft für meine Promotion. Diese Arbeit entstand während meiner vielseitigen Tätigkeit als wissenschaftlicher Mitarbeiter, Team- und Abteilungsleiter am Institut für Verkehrssystemtechnik des DLR e.V. Ich bedanke mich bei allen Menschen, die mich auf dem Weg unterstützt und inspiriert haben.

Für ihr Engagement und ihre Anregungen danke ich den Mitgliedern der Prüfungskommission Prof. Dr.-Ing. habil. Georg-Peter Ostermeyer, Prof. Dr.-Ing. Dr. h.c. mult. Eckehard Schnieder und Prof. Dr.-Ing. Karsten Lemmer. Als Doktorvater und Institutsleiter hat insbesondere Karsten Lemmer mit der für mich richtigen Mischung aus Führung und Freiraum zum Gelingen meiner Promotion beigetragen. Unsere Gespräche gaben mir stets wichtige Impulse und Motivation.

Ohne meine großartige Kollegin Katja Gutsche gäbe es vielleicht einzelne Forschungsergebnisse dieser Arbeit, doch erst durch ihre Überzeugungskraft und herausfordernde Art habe ich mich zur Promotion entschlossen. Ich danke ihr für den entscheidenden „Schubs“, die gemeinsame Forschung und die wichtigen Hinweise. Außerdem danke ich Christoph Lackhove, meinem langjährigen „Mitbüroener“ für vielfältige Diskussionen, Ansporn, unersetzliche Hinweise im Promotionsprozess und seine Freundschaft. Gleiches gilt für Benedikt Scheier, der zudem mit seinem tiefen Eisenbahnwissen, kritischem Blick und kreativen Ideen oft meine Perspektive erweitert hat. Ich bedanke mich gleichfalls bei meiner langjährigen Vorgesetzten Bärbel Jäger. Weit über diese Arbeit hinaus waren ihre Unterstützung, ihr Vertrauen und ihre Förderung für meine Entwicklung essenziell. Des Weiteren danke ich René Schenkendorf für die Impulse während unserer gemeinsamen Forschung. Hätte er mich nicht auf die Analysesoftware R aufmerksam gemacht, gäbe es einige sehr wichtige Grafiken in dieser Arbeit nicht. Auch den weiteren Mitgliedern des Life-Cycle-Management-Teams gilt mein Dank für viele Jahre vertrauensvoller, spannender und erfolgreicher Zusammenarbeit voller Spaß. Allen voran Florian Brinkmann, auf den ich mich immer verlassen konnte und der mir durch sein Engagement wichtige Freiräume verschaffte; Christian Lindner und Ikedichi Mbakwe, die die Forschung zur prädiktiven Instandhaltung komplementär zu meiner Arbeit vorantrieben und so unsere breiten Projekterfolge ermöglichten; Anja Bussmann, die mir eine wertvolle Diskussionspartnerin für alle ökonomischen Betrachtungen war.

Außerhalb des DLR bedanke ich mich bei Torsten Woicke, Bernd Schönberg und Wolfgang Knopf von der Deutschen Bahn für das Teilen ihres umfassenden Praxiswissens. Sie ermöglichten es mir, Felderfahrung bei der Weicheninstandhaltung zu sammeln, die ein unerlässlicher Baustein für die richtige Dateninterpretation ist.

Auf ganz besondere Weise danke ich meiner Lebensgefährtin Ina Lebedjew, die die Belastung der Promotion mitgetragen und dadurch für mich erleichtert hat. Sie hat auf unzählige Stunden Privatleben und gemeinsame Urlaube verzichtet, damit diese Arbeit entstehen konnte. Dies ist mit Nichts aufzuwiegen. Außerdem danke ich meinen Eltern für alles, was sie mir mitgaben, um dieses Leben zu führen.

Thomas Böhm

Inhaltsverzeichnis

Vorwort des Herausgebers	iii
Vorwort des Autors	v
Kurzfassung	xvii
Abstract	xix
1 Einleitung	1
1.1 Ausgangssituation und Motivation	1
1.2 Fokus, Zielstellung und Beitrag der Arbeit	2
1.3 Aufbau der Arbeit	6
2 Zustandsprädiktion von Eisenbahnweichen im technischen und wirtschaftlichen Kontext	9
2.1 Infrastrukturelement Weiche.....	9
2.1.1 Einordnung in das Gesamtsystem Eisenbahn	10
2.1.2 Aufbau und Funktion von Eisenbahnweichen	12
2.1.3 Umstellvorgang einer Eisenbahnweiche	17
2.2 Instandhaltung.....	18
2.2.1 Abnutzungsvorrat, verbleibende Nutzungsdauer und Instandhaltungsmaßnahmen	19
2.2.2 Lebenszyklus und Lebenszykluskosten.....	19
2.2.3 Instandhaltungskosten	21
2.2.4 Instandhaltungsstrategien	21
2.2.5 Schritte zur Umsetzung einer zustandsorientierten Instandhaltung	23
2.3 Zustandsorientierte Instandhaltung bei der Eisenbahninfrastruktur	24
2.3.1 Handlungsbedarf bei der Umsetzung einer zustandsorientierten Instandhaltung.....	25
2.3.2 Systeme zur Fernerfassung und Überwachung des Weichenzustands.....	26
2.4 Bestehende Prädiktionsmodelle und ihre Anwendbarkeit für Weichen.....	29
2.5 Zusammenfassung des Kapitels	39
3 Verfahren der Datenanalyse und der Zustandsprädiktion	41
3.1 Allgemeine Ansätze für die Zustandsbeurteilung	41
3.1.1 Ansätze des Prognose- und Zustandsmanagements.....	41
3.1.2 Datengetriebene Ansätze des Prognose- und Zustandsmanagements im Vergleich zur Wissensentdeckung in Datenbanken.....	43
3.2 Methoden für Prädiktionsmodelle.....	46

3.2.1	k-Nearest-Neighbor	46
3.2.2	Bayes'sche Netze	47
3.2.3	Künstliche Neuronale Netze	47
3.2.4	Entscheidungsbaumverfahren.....	48
3.2.5	Support Vector Machines.....	49
3.2.6	Vergleichende Bewertung der Verfahren.....	51
3.3	Metriken zur Qualitätsbeurteilung von Prädiktionsmodellen und Klassifikatoren .	52
3.3.1	Metriken bei der Klassifikation binärer Klassen	52
3.3.2	Metriken bei der Klassifikation multipler Klassen.....	55
3.3.3	Beurteilung bei unterschiedlichen Kosten einer falschen Klassifikation ...	59
3.4	Zusammenfassung des Kapitels	61
4	Anforderungen an das zu entwickelnde Vorhersagemodell des Weichenzustands	63
4.1	Allgemeine Anforderungen.....	63
4.2	Genauigkeitsanforderungen an das Prognosemodell aus wirtschaftlicher Sicht....	64
4.2.1	Transformation der stetigen Restnutzungsdauer auf eine binäre Klassifikation.....	65
4.2.2	Break Even Situation einer Vorhersage	67
4.2.3	Zusammensetzung der Kosten von Fehlalarmen und des Nutzens von verhinderten Störungen.....	71
4.3	Zusammenfassung des Kapitels	74
5	Datenquellen und Datenbasis für die Entwicklung des Prognosemodells	75
5.1	Analyse von Messdaten eines Weichendiagnosesystems und ihrer Eignung zur Zustandsbeurteilung	75
5.1.1	Entscheidung für SIDIS W als Weichendiagnosesystem.....	76
5.1.2	Zeitintervall der akquirierten SIDIS W Messdaten	76
5.1.3	Störungsdaten als Referenz für den Weichenzustand	78
5.1.4	Performance Evaluation des Weichendiagnosesystems SIDIS W unter technischen Gesichtspunkten.....	79
5.1.5	Performance Evaluation des Weichendiagnosesystems SIDIS W unter ökonomischen Gesichtspunkten	83
5.2	Analyse weiterer Einflussfaktoren bei der Zustandsbeurteilung	87
5.2.1	Detaillierte Beschreibung wesentlicher von SIDIS W gemessener Parameter	87
5.2.2	Klimabedingte Einflussfaktoren.....	87
5.2.3	Bauartbedingte Einflussfaktoren	93
5.3	Selektion der Rohdatenbasis zur Entwicklung des Prädiktionsmodells	102
5.4	Zusammenfassung des Kapitels	105
6	Prädiktionsmodell für den Zustand von Eisenbahnweichen	107
6.1	Fundamentales Konzept für die Prognose	107

6.1.1	Transformation der stetigen Restnutzungsdauer in multiple Klassen	108
6.1.2	Auswahl von Klassifikationsverfahren für die Prognose der Restnutzungsdauer und Metriken zu deren Bewertung	110
6.2	Experimenteller Vergleich der ausgewählten Klassifikationsverfahren	113
6.2.1	Individuelle Maximierung des Prognoseerfolgs der Klassifikationsverfahren anhand ihrer Parametrierung	115
6.2.2	Vergleich des Prognoseerfolgs bei bestmöglicher Parametrierung der Verfahren	119
6.2.3	Effekte alternativer RUL-Klassen auf den Prognoseerfolg	122
6.2.4	Performance Evaluation der Prognose unter ökonomischen Gesichtspunkten	124
6.3	Anwendung der Verfahren unter realen Bedingungen	126
6.4	Zusammenfassung des Kapitels	130
7	Zusammenfassung und Ausblick	133
7.1	Zusammenfassung der Ergebnisse	133
7.2	Ausblick auf weiterführende Forschung	136
Anhang	139
A	Mathematische Beschreibungen von Performance Metriken	139
A.1	Mathematische Beschreibung von Änderungen der Konfusionsmatrix	139
A.2	Formeln für die Berechnung ausgewählter Metriken	140
B	Ergänzende Informationen zur Datenbasis	143
B.1	Ergänzende Information zur Performance Evaluation von SIDIS W als Störungsdetektion	143
B.2	Bedeutung und Verwendung der Attribute in den verfügbaren Daten	145
B.3	Ergebnisse der Korrelationsanalyse zwischen SIDIS W Attributen und Klimadaten	149
B.4	Ergebnisse der Analyse zwischen SIDIS W Attributen und Bauarteigenschaften	153
B.5	Von der Zustandsreferenz ausgeschlossene Störungen	154
C	Ergänzende Informationen und Auswertungen zur Weichenzustandsprognose....	159
C.1	Prognoseergebnisse ausgewählter Verfahren bei bestmöglicher Parametrierung	159
C.2	Formeln zur Berechnung alternativer RUL-Klassen	163
C.3	Ergänzende Informationen zur Störungsvorhersage unter realen Bedingungen	164
Verwendete Abkürzungen	167
Verwendete Formelzeichen	171
Literaturverzeichnis	175

Abbildungsverzeichnis

Abbildung 1-1 :	Entwicklung der absoluten Verspätungsminuten (a) und in Relation zur Betriebsleistung (b) der DB Netz AG.....	2
Abbildung 1-2 :	Anteile verschiedener Gewerke an netzbedingten Verspätungsminuten der Jahre 2005-2010.....	3
Abbildung 2-1 :	Übersicht zur Einordnung von Weichenkomponenten in die Struktur des Systems Eisenbahn.....	11
Abbildung 2-2 :	Schematische Darstellung der Weichentypen.....	12
Abbildung 2-3 :	Konstruktionselemente einer einfachen Weiche in schematischer Darstellung.....	13
Abbildung 2-4 :	Beispiel eines einfachen Gleitstuhls (a) und eines Gleitstuhls mit Zungenrollern (b).....	14
Abbildung 2-5 :	Beispiel für einen innenliegenden (a) und einen neben dem Gleis liegenden Antrieb (b).....	14
Abbildung 2-6 :	Elektromechanischer Weichenantrieb S700K von Siemens (a) und elektrohydraulischer Antrieb L826H von Alcatel (b).....	15
Abbildung 2-7 :	Zeichnung eines Klammerverschlusses (a) und des Klinkenverschlusses CKA der Siemens Schweiz AG (b).....	16
Abbildung 2-8 :	Umstellvorgang einer einfachen Weiche mit Klammerverschluss.....	18
Abbildung 2-9 :	Phasen des Lebenszyklus.....	20
Abbildung 2-10 :	Verlauf des Abnutzungsvorrates und der Nutzungsdauer bei korrektiver Instandhaltung (a), präventiver Instandhaltung mit fixem Zeitintervall (b) und präventiver zustandsorientierte Instandhaltung (c).....	22
Abbildung 2-11 :	Prozess der Diagnose und der Prognose des Zustands.....	24
Abbildung 2-12 :	Entwicklung der Gesamtverspätungsminuten (a), der Störbestehenszeiten (b), der Gesamtstörungszahl (c) und der Instandhaltungsaufwände (d).....	26
Abbildung 2-13 :	Messkurvenverlauf der Diagnosesysteme SIDIS W (a) und Roadmaster 2000 Light (b) während der Umstellphasen einer Weiche.....	28
Abbildung 3-1 :	Vergleich des Prozesses bei der KDD und beim PHM.....	44
Abbildung 3-2 :	Beispiel der Klassifikation eines unbekannten Datensatzes mit kNN	46
Abbildung 3-3 :	Beispiel für den Aufbau eines Neurons und der Struktur eines aNN.....	48
Abbildung 3-4 :	Beispiel eines einfachen Entscheidungsbaumes.....	49
Abbildung 3-5 :	Beispiel für SVM mit verschiedenen Kernel-Funktionen.....	50
Abbildung 3-6 :	Beispiel für SVM-Transformation in einen höherdimensionalen Eigenschaftsraum mittels RBF-Kernel.....	50
Abbildung 3-7 :	Struktur der Konfusionsmatrix bei binärer Klassifikation.....	53
Abbildung 3-8 :	ROC-Graph mit zwei beispielhaften Klassifikatoren.....	54
Abbildung 3-9 :	ROC-Graph mit zwei beispielhaften Klassifikatoren bei unterschiedlichen Missklassifikationskosten.....	60

Abbildung 4-1 :	Reduktion der RUL-Prognose auf zwei Klassen auf Basis der Aussagen von Instandhaltern der DB Netz AG.....	67
Abbildung 4-2 :	Beispiel für die Visualisierung der Break Even Situation bei gegebenem Verhältnis von Kosten eines Fehlalarms zum Nutzen einer verhinderten Störung	70
Abbildung 4-3 :	Beispiel für die Visualisierung der Break Even Situation bei Verwendung eines besten und eines schlechtesten Kosten-Nutzen-Verhältnisses.....	71
Abbildung 5-1 :	Weichen und Zeitraum verfügbarer Messdaten aus SIDIS W	77
Abbildung 5-2 :	Prinzip der Alarme in SIDIS W am Beispiel der idealisierten Wirkleistungskurve.....	79
Abbildung 5-3 :	ROC-Graph Übersicht der verschiedenen Maßstäbe für die Zusammenfassung mehrerer Alarme	81
Abbildung 5-4 :	Visualisierung der Break Even Situation von SIDIS W, aufgelöst nach verschiedenen Störungshäufigkeiten je Umlauf (T_r); bei positivem Nutzen liegt SIDIS W links bzw. oberhalb der anderen Flächen.....	86
Abbildung 5-5 :	Beispiel eines von SIDIS W gemessenen Weichenumlaufs mit Kennzeichnung einiger wichtiger Attribute	88
Abbildung 5-6 :	Verlauf einiger Leistungswerte der Weiche 7 während des Aufzeichnungszeitraums	89
Abbildung 5-7 :	Arithmetisches Mittel aller linearen Korrelationskoeffizienten je Weiche und Umstellrichtung.....	91
Abbildung 5-8 :	Boxplot der Korrelationskoeffizienten von SIDIS W Attributen zur Temperatur und zur Luftfeuchtigkeit.....	92
Abbildung 5-9 :	Dendrogramm der Weichen und ihrer Bauarteigenschaften	97
Abbildung 5-10 :	MCC der Klassifikation der Bauartclusterzugehörigkeit nach SIDIS W Messdaten	99
Abbildung 5-11 :	MCC der Klassifikation der Bauartclusterzugehörigkeit je Clusteranzahl	100
Abbildung 5-12 :	Gegenüberstellung der MCC-Werte der Zufallscluster- und der Bauartclusterzugehörigkeit	101
Abbildung 6-1 :	Ergebnis eines ersten Prognoseversuchs mit zufällig gewählten Parametern für die vier Klassifikationsverfahren	114
Abbildung 6-2 :	Prädiktionserfolg von kNN bei verschiedenen k-Werten	116
Abbildung 6-3 :	Prädiktionserfolg bei unterschiedlicher Anzahl von verdeckten Schichten und Neuronen im aNN.....	117
Abbildung 6-4 :	Prädiktionserfolg bei verschiedenen Teilungskriterien im DTree, jeweils mit und ohne Beschneiden	118
Abbildung 6-5 :	Prädiktionserfolg von SVM mit RBF Kernel und verschiedenen Einstellungen	119
Abbildung 6-6 :	Konfusionsmatrix der Prognose mit kNN als Anteil an der tatsächlichen RUL-Klasse.....	120
Abbildung 6-7 :	Konfusionsmatrix der Prognose mit DTree als Anteil an der tatsächlichen RUL-Klasse	121
Abbildung 6-8 :	Konfusionsmatrix der Prognose mit SVM als Anteil an der tatsächlichen RUL-Klasse	121

Abbildung 6-9 :	Visualisierung der Break Even Situation der Störungsprognose, Auflösung im Intervall FP Rate bis 0.2 und wenn bei jedem 1000. Umlauf oder seltener eine Störung eintritt (T_r); bei positivem Nutzen liegen die farbigen Flächen links bzw. oberhalb der anderen Flächen.....	125
Abbildung 6-10 :	ROC-Graph der paarweisen Vorhersage unter realen Bedingungen von je einer Störung anhand jeder einzelnen zurückliegenden Störung	129
Abbildung A-1 :	ROC-Kurve von zwei beispielhaften Klassifikatoren	141
Abbildung B-1 :	ROC-Graph der verschiedenen Maßstäbe für die Zusammenfassung mehrerer Alarme.....	144
Abbildung B-2 :	ROC-Graph der verschiedenen Maßstäbe vor und nach der korrelationsgemäßen Temperaturanpassung	152
Abbildung B-3 :	Violinplot der SIDIS W Attribute Pleer (a) und Pverschl (b).....	153
Abbildung B-4 :	ROC-Graph der verschiedenen Maßstäbe vor und nach dem Ausschluss von Störungen.....	158
Abbildung C-1 :	Verteilung der MCC-Werte einer paarweisen Vorhersage unter realen Bedingungen von je einer Störung anhand jeder einzelnen zurückliegenden Störung	165
Abbildung C-2 :	Histogramm der MCC-Werte bei der Störungsprognose je Weiche mit und ohne Bauartzuordnung	165

Tabellenverzeichnis

Tabelle 2-1 : Eigenschaften der Weichendiagnosesysteme SIDIS W und Roadmaster 2000 Light	29
Tabelle 2-2 : Zusammenfassung bestehender Prädiktionsmodelle für Weichen und Bewertung ihrer Anwendbarkeit	38
Tabelle 3-1 : Zusammenfassung von Eigenschaften ausgewählter Methoden für Prädiktionsmodell.....	51
Tabelle 3-2 : Beispiel 1 einer Konfusionsmatrix	53
Tabelle 3-3 : Beispiel 2 einer Konfusionsmatrix	53
Tabelle 3-4 : Konfusionsmatrizen der Klassifikatoren aus Abbildung 3-8.....	54
Tabelle 3-5 : Struktur der Konfusionsmatrix bei multipler Klassifikation	56
Tabelle 3-6 : Konfusionsmatrix bei multipler Klassifikation mit der ersten Klasse als positive Klasse	56
Tabelle 3-7 : Konfusionsmatrix bei multipler Klassifikation mit der letzten Klasse als positive Klasse	56
Tabelle 3-8 : Zusammenfassung der Eigenschaften verschiedenen Performance Metriken .	58
Tabelle 3-9 : Konfusionsmatrizen der Klassifikatoren und Kostenmatrizen der unterschiedlichen Missklassifikationskosten aus Abbildung 3-9	60
Tabelle 4-1 : Auflistung aller Anforderungen an die zu entwickelnde Prognose	73
Tabelle 5-1 : Ausgewählte Attribute des SAP-Auszugs von Störungsmeldungen und ihre Bedeutung	78
Tabelle 5-2 : Gegenüberstellung von Alarmen in SIDIS W mit den tatsächlichen Störungen	80
Tabelle 5-3 : Werte der FP und TP Rate bei verschiedenen Maßstäben für die Zusammenfassung mehrerer Alarme	82
Tabelle 5-4 : Positionen der Kosten eines Fehlalarms und des Nutzens einer richtig prognostizierten Störung	85
Tabelle 5-5 : Ausgewählte Attribute, die in SIDIS W aufgezeichnet werden und deren Bedeutung	88
Tabelle 5-6 : Attribute der Wetterstation und ihre Bedeutung	90
Tabelle 5-7 : Bauarteigenschaften der 29 Weichen	95
Tabelle 5-8 : Häufigkeit von Kombinationen mit einer bestimmten Attributanzahl (links) und eines bestimmten Attributs (rechts)	99
Tabelle 5-9 : Status aller Anforderungen an die zu entwickelnde Prognose	106
Tabelle 6-1 : Klassen der RUL mit ihren jeweiligen Intervallen in Tagen	109
Tabelle 6-2 : TP Rate je Klasse im Vergleich der Verfahren	120
Tabelle 6-3 : Vergleich der ähnlichkeitsgewichteten, makroskopischen AUC je Klassenstruktur und Verfahren	123
Tabelle 6-4 : Vergleich der Prognoseresultate je Verfahren, wenn Daten vor einem bestimmten Zeitpunkt für die Vorhersage ab diesem Datum genutzt werden ..	127
Tabelle 6-5 : Anzahl der nach Ursache und Bauteilgruppe unikaten Störungen (Referenzen) je Bauartcluster	128

Tabelle 6-6 : Status aller Anforderungen an die zu entwickelnde Prognose	131
Tabelle A-1 : Tatsächliche Klasse der Objekte und ihr Zugehörigkeitswert zu positiv	141
Tabelle B-1 : Verfügbare Attribute in SIDIS W, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell.....	147
Tabelle B-2 : Verfügbare Attribute des SAP-Auszugs dokumentierter Störungen an Weichen, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell	148
Tabelle B-3 : Verfügbare Attribute der Klimadaten, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell	149
Tabelle B-4 : Standardabweichung für das arithmetische Mittel aller linearen Korrelationskoeffizienten je Weiche und Umstellrichtung	150
Tabelle B-5 : SAP-Störungseinträge und die Begründung für ihren Ausschluss von der Zustandsreferenz bei der Prognosemodellentwicklung	157
Tabelle C-1 : Konfusionsmatrix der Prognose mittels kNN bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen.....	160
Tabelle C-2 : Konfusionsmatrix der Prognose mittels DTree bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen	161
Tabelle C-3 : Konfusionsmatrix der Prognose mittels SVM bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen.....	162

Kurzfassung

Im Schienenverkehr verursachen Verspätungen jährlich einen enormen volkswirtschaftlichen Schaden. Eisenbahnunternehmen leiden unter den negativen Folgen für ihr Image sowie unter den Kosten der Verspätungen und der Instandhaltung zur Behebung von Störungen. Eine der Hauptursachen für infrastrukturbedingte Verspätungen sind Weichenstörungen. Ihre negativen Auswirkungen ließen sich verhindern, wenn die Störung rechtzeitig bekannt wäre, um effektive und effiziente Gegenmaßnahmen zu planen und auszuführen. Im Rahmen dieser Arbeit wird ein datengetriebener Ansatz zur präzisen und verlässlichen Prognose von Weichenstörungen in Form der verbleibenden Nutzungsdauer (engl. Remaining Useful Life) RUL vorgestellt.

Ausgangsbasis sind Stellstromdaten aus dem zweijährigen Realbetrieb von 29 Weichen der DB AG, die durch ein kommerzielles Weichendiagnosesystem rückwirkungsfrei und kontinuierlich am Weichenantrieb erfasst wurden. Die Analyse dieser Daten zeigt, dass das kommerzielle System allein nicht für die Prognose ausreicht. Daher wird der Einfluss weiterer Parameter strukturiert untersucht. Insbesondere die Temperatur und die Luftfeuchtigkeit spielen für die Interpretation der Stellströme eine wesentliche Rolle. Auch die Bauartigenschaften der Weichen sind entscheidend. Ihr Einfluss wird mit einer eigenen, neuartigen Methode analysiert, da die Voraussetzungen für herkömmliche Analyseverfahren nicht gegeben sind. Die entwickelte Heuristik kombiniert Clustering und Klassifikation, um jene Bauartigenschaften zu identifizieren, die das unterschiedliche Weichenverhalten bestimmen. Statt Weichen individuell zu behandeln, wird die Prognose auf Bauartgruppen ausgerichtet, weil Weichen innerhalb der Gruppen ähnliche Stellstromverläufe besitzen. Auf diese Weise kann auch der Zustand neuer Weichen prognostiziert werden, für die noch keine historischen Daten vorliegen.

Auf der Grundlage der analysierten Zusammenhänge entsteht eine hochdimensionale Rohdatenbasis heterogener Quellen. Sie werden den dokumentierten tatsächlichen Störungen als Zustandsreferenz gegenübergestellt, womit für jedes Messtupel die RUL gegeben ist. Der Prognose liegt die Idee zugrunde, die stetige RUL in diskrete Klassen zu transformieren, die jeweils ein bestimmtes Intervall an RUL-Zeiten zusammenfassen. Anschließend ist eine Abbildung der Messtupel auf die Klassen zu generieren, so dass unbekannte Messtupel einer RUL-Klasse zugeordnet werden können. Die Prognose wird zum Klassifikationsproblem multipler Klassen, für das zur Daten- und Problemstruktur passende Klassifikationsverfahren gesucht und anhand geeigneter Metriken verglichen werden. Weil die Performance der Verfahren stark von ihrer Parametrierung abhängt, werden jene Einstellungen herausgearbeitet, bei denen der Erfolg maximal ist. Außerdem wird der Effekt verschiedener Klassenstrukturen untersucht. So entstehen zusätzlich Grundregeln für die Klassenstruktur, die die Anforderungen an Planbarkeit, Granularität und Genauigkeit der RUL-Prognose berücksichtigen.

Im Ergebnis weist die Arbeit nach, dass zwei Verfahren Weichenstörungen rechtzeitig, präzise und verlässlich prognostizieren und zudem wirtschaftlich arbeiten. Mit einer selbst entwickelten Methode wird bewiesen, dass die Genauigkeit in verschiedenen Situationen hoch genug ist, damit durch richtige Vorhersagen höhere Verspätungskosten eingespart als durch Fehlalarme unnötige Instandhaltungskosten induziert werden. So könnte ein Eisenbahnunternehmen mindestens 50 Prozent der resultierenden Verspätungen verhindern und gleichzeitig Kosten senken.

Abstract

Delays in the rail transport induce a high national economic loss every year. Railway companies suffer from the negative reputation as well as from the delay and maintenance costs resulting from failures. One of the main causes for infrastructure related delays are switch failures. Their negative effects could be prevented, if a failure would be revealed early enough to schedule and execute effective and efficient maintenance measures. The thesis at hand presents a data driven approach for a precise and reliable prediction of switch failures by providing the Remaining Useful Life (RUL).

The work is based on data of the electric power consumed by the motor during switch repositioning. Real life data of 29 switches in the DB AG network has been recorded continuously with a nonreactive commercial switch diagnostic system over a two year period. Analysing this source material, it is shown that the commercial system alone is not sufficient to predict or even diagnose failures. Therefore, the influence of additional parameters is investigated. In particular, the temperature and the humidity play an important role in interpreting the power measurements. Also, switch construction characteristics are essential. Because the data structure does not fulfil the mathematical requirements to use conventional techniques a novel method is developed to qualify the dependencies between construction, power measurements and condition. The heuristic combines clustering and classification in order to identify characteristics determining different switch behaviour. Instead of treating each switch individually, this enables a prediction according to a certain group of switches because switches within a group have very similar power measurement patterns. By using this finding, it is possible to also forecast new switches for which no historical data set is available.

Based on the analysed interrelations, a high dimensional data collection of heterogeneous sources is assembled. It is compared to the documented actual switch failures as ground truth for the condition reference, hence allowing to assert the RUL for each measurement tuple. The underlying idea of the prediction is to transform the continuous RUL into discrete classes, which summarise a certain RUL interval. Subsequently, a mapping of measurement tuples to the classes is generated, so that unknown tuples can be assigned to a RUL-class. The prediction becomes a classification problem of multiple classes, for which classification techniques have to be derived and evaluated by metrics that suit the data structure as well as the problem structure. Since the performance of the eligible techniques is heavily depending on their parametrisation, the thesis investigates those settings allowing for the most forecast success. Furthermore, the effects of different RUL-class structures are analysed. This provides basic rules for the RUL-ranges that meet requirements of schedulability, granularity and precision for the prediction.

The thesis proofs that two classification techniques are able to predict switch failures early enough, precisely and reliable and are also economically efficient. The latter is verified by a self-developed method. It shows for a wide range of scenarios that the precision is high enough to save more delay costs by correctly predicted failures than unnecessary maintenance costs are induced by false alerts. Therefore, the presented solution allows a railway company to prevent at least 50 percent of the delay time resulting from switch failures and at the same time reduce costs.

1 Einleitung

Im Schienenverkehr verursachen Verspätungen jährlich einen enormen volkswirtschaftlichen Schaden in kaum bezifferbarer Höhe. Gleichzeitig zählt Unpünktlichkeit zu den größten Ärgernissen für Fahrgäste. Davon zeugen zahlreiche, überwiegend negative Berichte von Reisenden, beispielsweise unter dem Titel „*Gute Bahn - Schlechte Bahn*“ auf den Internetseiten des Verkehrsclubs Deutschland [VCD12]. Eisenbahnunternehmen, allen voran die Deutsche Bahn Aktiengesellschaft (DB AG), leiden nicht nur unter den negativen Folgen für ihr Image, sondern auch unter den Kosten der Verspätungen und der Instandhaltung zur Behebung von Störungen. Die negativen Auswirkungen einer infrastrukturbedingten Störung ließen sich nur verhindern, wenn diese rechtzeitig bekannt wäre, um effektive und effiziente Gegenmaßnahmen zu planen und auszuführen. Solch eine Prognose existiert jedoch bisher nicht valide, sie muss erst erforscht werden. Daher ist das Ziel dieser Arbeit eine präzise und verlässliche Vorhersage von Weichenstörungen, die zudem die Instandhaltungskosten senkt.

1.1 Ausgangssituation und Motivation

Die Senkung der Instandhaltungskosten und die Erhöhung der Pünktlichkeit sind für die Eisenbahninfrastrukturbetreiber der Europäischen Union (EU) seit vielen Jahren wichtige und nach wie vor aktuelle Ziele. So konstatierte etwa der Internationale Eisenbahnverband UIC schon 2008 in seinem kontinuierlichen Benchmarking der Infrastrukturkosten: *„Die Verbesserungen bei den Instandhaltungskosten waren in Europa in den letzten Jahren durchschnittlich bescheiden.“* [UIC08, S. 9] Bei Europas größtem Infrastrukturbetreiber DB Netz AG sind die Instandhaltungskosten in den letzten Jahren sogar gestiegen [DB 14a]. Gleichzeitig verringerte sich die Pünktlichkeit. Dies lässt sich anhand der Infrastrukturzustands- und Netzentwicklungsberichte nachvollziehen, die seit 2011 jährlich veröffentlicht werden. Die Summe der Verspätungsminuten (VMin) ist 2015 mit 174.6 Mio. im Vergleich zu 145.2 Mio. in 2008 deutlich gestiegen [DB 11, DB 14b, DB 16], siehe Abbildung 1-1(a). In dieser Summe sind Verspätungen mit sehr unterschiedlichen Ursachen enthalten, beispielsweise höhere Gewalt, Personen oder Tiere im Gleis, Primärverspätungen durch Verkehrsunternehmen, aber auch durch Netzinfrastukturstörungen [Gün10]. Gerade letztere, die netzbedingten Verspätungsursachen, sollten jedoch durch den technologischen Fortschritt bei der Störungsdiagnose und -prognose deutlich gesunken sein [Fri04, S. 27]. Dass dem nicht so ist, zeigen die Zahlen in Abbildung 1-1(b). In Relation zur Betriebsleistung auf den Strecken der DB Netz AG haben sich die netzbedingten VMin nur leicht verringert.

Dabei setzt die DB Netz AG an vielen verschiedenen Stellen im Netz Diagnosesysteme mit der Absicht ein, den Zustand der Infrastruktur zu überwachen und Ausfälle prognostizieren zu können. Diese Systeme erfüllen jedoch nicht die Erwartungen des Betreibers.

„Zumeist stellen die Diagnosen im Störfall dem Instandhaltungspersonal Zustandsdaten der Anlage zur Verfügung, um die Entstörung zu beschleunigen. Der gestörte Zustand der Anlage soll aber gar nicht eintreten und das eigentliche Ziel der Zustandsüberwachung und einer frühzeitigen Meldung wird nur selten erfüllt.“ [Hol12, S. 20]

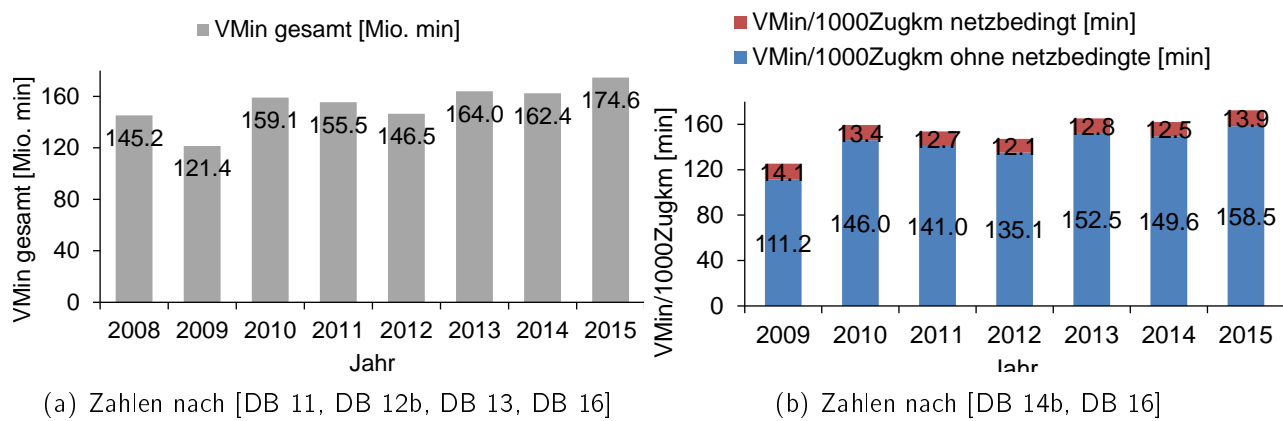


Abbildung 1-1: Entwicklung der absoluten Verspätungsminuten (a) und in Relation zur Betriebsleistung (b) der DB Netz AG

Zusammenfassend betrachtet sind die von den Eisenbahninfrastrukturbetreibern selbst gesteckten Ziele nicht erreicht worden, da sich Instandhaltungskosten und Verspätungsminuten kaum verringert haben. Die Anwendung existierender Systeme ist für die Zustandsprädiktion nicht ausreichend weit entwickelt. Aus dieser Situation leitet sich die Motivation zu der vorliegenden Arbeit ab. Mit ihrem Inhalt soll ein wesentlicher Beitrag geleistet werden, Verfahren und Systeme der Zustandsprognose zu entwickeln – für eine leistungsstärkere, wettbewerbsfähigere Schieneninfrastruktur, mehr Pünktlichkeit und zufriedenere Reisende.

1.2 Fokus, Zielstellung und Beitrag der Arbeit

Eine Zustandsprädiktion setzt voraus, dass der Infrastrukturzustand kontinuierlich erfasst und überwacht wird. Als kontinuierlich wird hier im Sinne der VDI 2888 [VDI12] ein Erfassungsintervall verstanden, dass jede nutzungsbedingte Zustandsveränderung unmittelbar offenlegt. Bei seiner Größe und der hohen Zahl an Infrastrukturelementen ist eine vollständige Überwachung des Eisenbahnnetzes jedoch unwirtschaftlich. Daher konzentriert sich diese Forschungsarbeit auf jene Elemente, deren Störungen für einen hohen Anteil an Verspätungsminuten verantwortlich sind, ein erhöhtes Sicherheitsrisiko für den Eisenbahnbetrieb darstellen und/oder deren Behebung außerhalb der Betriebspausen besonders risikoreich für den Instandhalter vor Ort ist.

Wie die Ergebnisse des Projektes InfraCost aus dem Jahr 2002 belegen, ist die Störanfälligkeit von Weichen ein Problem für viele Europäische Eisenbahnen. „Broken down into infrastructure components again, switch and signalling installations tend to account for the highest share of delays.“ [BR02, S. 35] Wie in Abbildung 1-2 zu sehen ist, sind bei der DB Netz AG die signaltechnischen Anlagen mit 33 Prozent und die Weichen und Kreuzungen mit 19 Prozent die beiden Gewerke mit dem höchsten Anteil an VMin [Rau10]. Während unter den signaltechnischen Anlagen Stellwerke, Signale, Achszähler, Gleisstromkreise und Balisen zu einer heterogenen Gruppe zusammengefasst sind, bilden die Weichen und Kreuzungen eine eher homogene Gruppe von Elementen und sind damit ein Störungsschwerpunkt.

Weichen sind zwar ein störanfälliges Element, aber im betrieblichen Sinn essentiell. Sie erlauben im Zugverkehr den Gleis- bzw. Richtungswechsel ohne Halt. Die Störung einer Weiche betrifft fast immer mehr als ein Gleis. Darüber hinaus sind Weichen für die Instandhalter mitunter schwer

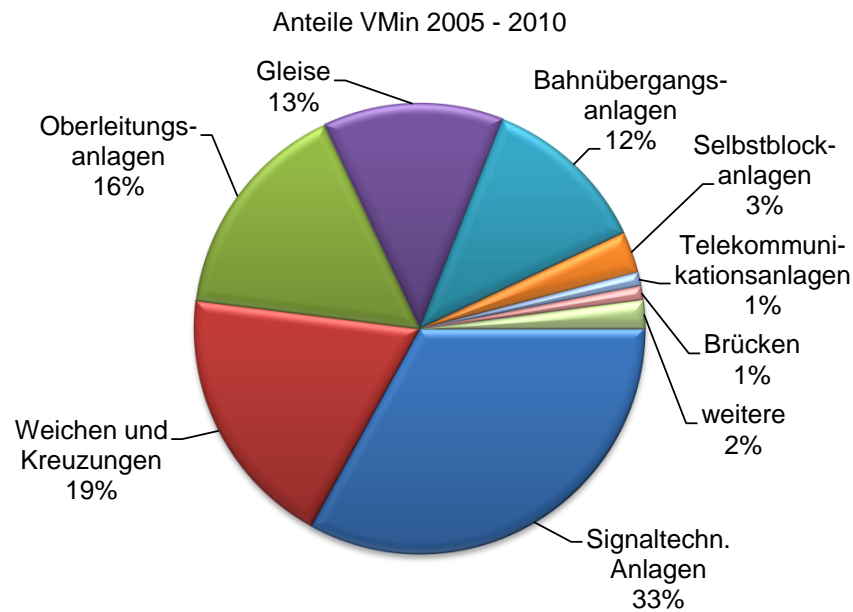


Abbildung 1-2: Anteile verschiedener Gewerke an netzbedingten Verspätungsminuten der Jahre 2005-2010 (in Anlehnung an [Rau10])

erreichbar, so dass allein die Anfahrt bei einer Störung zeitaufwendig ist. Zusätzlich stellen die Instandhaltungsarbeiten im Gleisbett ein erhöhtes Risiko für das Personal dar. Immer wieder kommt es dabei zu Unfällen mit Verletzten oder auch Toten [Ber10].

Weichen sind damit prädestiniert für eine kontinuierliche Zustandsüberwachung sowie eine darauf aufbauende, automatische Diagnose und Prognose. Während für die Diagnose, also die Beurteilung der Istsituation und der Störungsursache, in den vergangenen Jahren viele Lösungsansätze präsentiert wurden (bspw. von Zhou et al. [ZAB⁺02], García Márquez et al. [GMWR07], Chamroukhi et al. [CSAA08], Silmon [Sil09], Atamuradov et al. [ACBS09], Bai [Bai10], Eker et al. [ECK10] oder Asada et al. [ARK13]), mangelt es an Lösungen für die Prädiktion. Doch erst die Vorhersage von Störungen hat das Potenzial, die Verspätungen und die Instandhaltungskosten deutlich zu senken. Aus diesem Grund und wegen der besonderen Bedeutung von Weichen, fokussiert die vorliegende Arbeit die Entwicklung eines Zustandsprognoseverfahrens für Eisenbahnweichen.

Modellgetriebene Methoden ermöglichen zwar üblicherweise genauere Prognosen, aber ihre Erstellung ist umso schwerer und zeitaufwendiger, je komplexer und komplizierter das zu prognostizierende System ist [UGL08]. Die Studie EcoSwitch hat 75 Einflussfaktoren auf den Weichenzustand identifiziert [JZ02], so dass ein modellgetriebener Ansatz als weniger aussichtsreich erscheint als ein datengetriebener Ansatz. Als Instrument für die Überwachung können kommerziell erhältliche und bereits im Feldeinsatz befindliche Weichendiagnosesysteme dienen. Auch wenn diese Systeme bisher deutliche Schwächen bei der Verlässlichkeit aufweisen [Böh12b] und per se keine Prädiktion ermöglichen [Hol12], liefern sie doch umfassende Verhaltensdaten über einen längeren Zeitraum. Diese bilden die Grundlage für einen datengetriebenen Ansatz bei der Entwicklung eines Prognoseverfahrens.

In Anbetracht der erwähnten Vielzahl von Einflussfaktoren auf den Weichenzustand sowie in Abhängigkeit von der Anzahl an Weichen und Weichenumläufen, wird die Datenbasis nicht nur

hochdimensional, sondern auch sehr umfangreich. Würden beispielsweise Überwachungsdaten aus einem Jahr von 10 Weichen, die einmal stündlich umlaufen, in eine Tabelle eingetragen, so ergäben sich bereits 87600 Einträge. Das Netz der DB AG umfasst allerdings 67071 Weichen und Kreuzungen [DB 16], von denen viele mehrmals in der Stunde umlaufen, besonders an Punkten mit hoher Verkehrsdichte. Bei dem zu erwartenden Datenumfang setzt die vorliegende Arbeit auf Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD) mit ihren modernen Analyseverfahren wie zum Beispiel Data Mining.

Mit dem gesetzten Fokus – durch einen datengetriebenen Ansatz mittels Data Mining ein Zustandsprognoseverfahren für Eisenbahnweichen zu entwickeln – sollen folgende Forschungsfragen beantwortet werden:

1. Wie kann bei diesem Ansatz ein Verfahren gestaltet werden, mit dem sich mindestens **50 Prozent der Weichenstörungen rechtzeitig vorhersagen** lassen, so dass ausreichend Zeit besteht, eine Instandhaltung zu planen und durchzuführen?
2. Wird der Netzbetreiber mit dem Verfahren in die Lage versetzt, auch **50 Prozent der durch Weichen verursachten Verspätungsminuten zu reduzieren**? Dies würde die Pünktlichkeit um 1.34 Mio. VMin oder rund 2.5 Jahre verbessern.
3. Arbeitet das Verfahren wirtschaftlich, trägt also die Lösung auch zur **Senkung der Instandhaltungskosten** bei?
4. Können die Überwachungsdaten, auf denen die Prognose beruht, mit einer **minimalen Sensorausrüstung** gemessen werden, die **vollständig rückwirkungsfrei auf die Leit- und Sicherungstechnik** ist? Denn dadurch wäre sichergestellt, dass der Netzbetreiber bestehende Anlagen nachrüsten kann, ohne dass eine Zulassung durch das Eisenbahn-Bundesamt (EBA) notwendig wird.

Die vorliegende Arbeit liefert zahlreiche Beiträge und neue Ansätze, um diese und damit verbundene Fragen zu beantworten. Dazu zählt unter anderem die Klärung, was der Begriff „*rechtzeitig*“ im Zusammenhang mit dem Bekanntwerden einer Störung bedeutet. Diese Frage ist in bisherigen, wissenschaftlichen Arbeiten nicht gestellt oder beantwortet worden, wahrscheinlich weil existierende Ansätze meist nur die Diagnose fokussieren. In dieser Arbeit wird nicht nur erstmals gezeigt, wie die Prognose in Form einer verbleibenden Nutzungsdauer (engl. Remaining Useful Life) (RUL) erfolgen kann. Sie liefert zudem Grundregeln zum Gestalten der RUL-Aussagen, die die Planbarkeit, die Präzision, die Verlässlichkeit und die Nutzerakzeptanz berücksichtigen. Denn die Beurteilung des Zustands ist schwankungsanfällig, gerade bei längerfristigen Prognosen. Gibt eine Störungsvorhersage mal mehr und mal weniger Zeit bis zur Störung aus, wirkt sie auf den Nutzer unzuverlässig, selbst wenn die jeweilige Prognose unter gleichbleibenden Zustandseinflüssen korrekt ist.

Für die eigentliche Prognose stehen zahlreiche Verfahren zur Auswahl, beispielsweise künstliche Neuronale Netze oder Support Vector Machines. Doch welche Kandidaten sind für die vorliegende Problemstellung geeignet und bei welcher Parametrierung liefern sie ausreichend gute Ergebnisse? Dies mittels systematischer Analyse und Bewertung zu beantworten, ist ein Schwerpunkt der Arbeit. Gerade die Parametrierung ist dabei zentral, da diese maßgeblich den Prognoseerfolg determiniert. Denn mitunter erzeugt ein und dasselbe Verfahren zwei gegensätzliche Ergebnisse.

Je nach Parametereinstellung kann dabei die Vorhersage komplett misslingen oder zu fast 100 Prozent korrekt sein. Da kein ultimativ bestes Verfahren und keine immer passende Parametrierung existieren, bedarf es bei jeder neuen, mittels Data Mining zu lösenden Problemstellung einer experimentellen Herangehensweise. Diese systematische Suche nach passenden Verfahren und ihren bestmöglichen Einstellungen ist enorm zeitaufwendig, aber ein wichtiger Forschungsbeitrag, um unzählige Data Mining Probleme in der Biologie, der Medizin, der Wirtschaft, etc. zu lösen. Das gilt insbesondere für die vorliegende Arbeit zur Weichenzustandsprognose, weil das dafür notwendige mathematische, daten- und implementierungstechnische Know-how nicht Schwerpunkt eines Infrastrukturbetreibers ist.

Die Frage, ob das zu entwickelnde Prognoseverfahren wirtschaftlich arbeitet, muss immer in Verbindung mit den Einsatzbedingungen beantwortet werden. Je nach Störungshäufigkeit, Kosten eines Fehlalarms und dem Nutzen aus einer verhinderten Störung ergeben sich unterschiedliche Anforderungen an die Genauigkeit. Diese Arbeit beinhaltet eine eigens entwickelte Methode, um nachzuweisen und auch zu visualisieren, unter welchen Bedingungen die Prognose ausreichend genau ist, damit sie die Instandhaltungskosten senkt. Die Methode definiert die Break Even Situation (in Anlehnung an den Break Even Point bei der Produktion), in der sich eine Vorhersage im ökonomischen Gleichgewicht befindet. D.h. das Prädiktionsverfahren spart durch richtig vorhergesagte Störungen genau so viel Kosten ein, wie es durch Fehlalarme unnötigerweise verursacht. Diese Methode ist dabei nicht auf Weichen beschränkt, sondern für eine ökonomische Performanceevaluation von Prognosen und Klassifikationen allgemein anwendbar.

Die Störungsprognose soll auf Daten von Sensoren beruhen, die rückwirkungsfrei und für den Einsatz im Eisenbahnumfeld zugelassen sind. Solche Sensoren sind zum Beispiel Bestandteil der bestehenden Weichendiagnosesysteme mit ihren Stellstrommessungen. Die vorliegende Arbeit analysiert, inwieweit solch ein System bereits für die Vorhersage genutzt werden kann. Dazu wird auch der neue Ansatz genutzt, Grenzwertalarme aus Stellstromabweichungen in unterschiedlicher Weise zu kombinieren. Den zweiten Schwerpunkt setzt diese Arbeit auf die Analyse, welche Daten überhaupt notwendig sind, damit eine Störungsprädiktion gelingt. Hierzu wurden erstmals die klimabedingten und die bauartbedingten Zusammenhänge zu Stellstrommessungen quantifiziert bzw. identifiziert. Die Bauarteigenschaften sind deshalb relevant, weil die Stellstromwerte je Weiche sehr verschieden sein können. Wird nun dieses unterschiedliche Verhalten mit Bauarteigenschaften in Verbindung gebracht, lässt sich auch der Zustand unbekannter Weichen prognostizieren. Der Zusammenhang wird dabei mit einer neuartigen Analysemethode bewiesen, die zwei Schwierigkeiten überwindet, an denen Standardanalysemethoden wie beispielsweise die Korrelationsanalyse, Hotellings T^2 -Test oder die Varianzanalyse (ANOVA) scheitern. Zum einen wird der Zusammenhang von umfangreichen, kontinuierlich erhobenen und schwankenden Stellstromwerten zu wenigen statischen Eigenschaften wie etwa dem Weichentyp, dem Radius oder der Schwellenart hergestellt. Zum anderen werden nicht normalverteilte Werte mit ungleichen Varianzen analysiert.

Insgesamt betritt diese Arbeit damit Neuland in Bezug auf die Analyseverfahren zur Identifikation notwendiger Daten für die Weichenzustandsprognose, den Umfang und die Realitätsnähe der Datenbasis, die präzise und verlässliche Vorhersage in Form einer Restnutzungsdauer sowie den Nachweis einer wirtschaftlichen Prognosegenauigkeit. Die entwickelten Methoden und gewonnenen Erkenntnisse sind dabei auch auf andere Systeme übertragbar.

1.3 Aufbau der Arbeit

Nach dieser Einleitung stellt Kapitel 2 die Thematik der Störungsvorhersage in einen breiteren Kontext, um das grundlegende Verständnis zu erleichtern. Dazu wird die Weiche in das Gesamtsystem Eisenbahn eingeordnet sowie ihr Aufbau und ihre Funktion im Detail erläutert. So können die im Verlauf der Arbeit getroffenen Aussagen und Schlussfolgerungen besser nachvollzogen werden. Zudem wird allgemein auf die Vorhersage als Bestandteil einer wirtschaftlichen Instandhaltungsstrategie eingegangen. Die spezielle Situation zur zustandsorientierten Instandhaltung bei der DB Netz AG wird ebenfalls erläutert, wodurch insbesondere der Handlungsbedarf konkretisiert wird. Detailliert diskutiert das Kapitel bisherige Ansätze bzw. Arbeiten als Stand der Wissenschaft, die eine Zustandsprädiktion von Weichen beinhalten.

In Kapitel 3 folgen Grundlagen zu Verfahren der Datenanalyse und der Zustandsprädiktion, um einen Überblick über mögliche Herangehensweisen zu geben. Neben einzelnen Verfahrensgruppen zur Prognose werden auch Metriken zur Erfolgsmessung beschrieben und bewertet.

Aus der Diskussion zu bestehenden Ansätzen leitet anschließend das Kapitel 4 konkrete Anforderungen an die zu entwickelnde Prädiktion ab. Neben allgemeinen Anforderungen steht die Genauigkeitsanforderung aus wirtschaftlicher Sicht im Mittelpunkt. Hier wird die Methode zur Break Even Situation entwickelt. Im Laufe der Arbeit werden die definierten Anforderungen dazu herangezogen, die einzelne Schritte und schließlich das Gesamtergebnis zu beurteilen.

Kapitel 5 beschäftigt sich mit der Selektion und Vorverarbeitung von Daten als Grundlage der Prognose. Es werden die als Zustandsreferenz genutzten tatsächlichen Störungen an Weichen beschrieben. Mit ihrer Hilfe wird die Performance eines herkömmlichen Weichendiagnosesystems evaluiert, aus technischer und aus ökonomischer Sicht. Letzteres geschieht mit der Methode zur Break Even Situation. Aus der Evaluation geht hervor, ob die originären Messdaten eines solchen Diagnosesystems bereits ausreichen, um die gestellten Anforderungen zu erfüllen. Daraufhin werden erfolgsentscheidende Attribute von Stellstrommessungen, Klimadaten und Bauarteigenschaften bestimmt und miteinander kombiniert. Dafür müssen zum Teil neuartige Analyseverfahren entwickelt und angewendet werden. So entsteht jene Datenbasis, die für die Prognose unverzichtbar ist.

Unter Nutzung der Datenbasis setzt Kapitel 6 die Prognose um. Es stellt zunächst die fundamentale Idee zur Vorhersage der verbleibende Nutzungsdauer (engl. Remaining Useful Life) (RUL) vor. Danach werden konkrete Data Mining Verfahren und Metriken ausgewählt oder erstellt, die bei der vorliegenden Problemstruktur und Datenbasis anwendbar sind. Der Prognoseerfolg verschiedener Verfahren wird anhand ihrer Parametereinstellungen maximiert, wodurch der gemeinsame Vergleich möglich wird. Dieser weist nach, dass die entwickelte Prognose präzise, verlässlich und wirtschaftlich ist. Zudem wird dargestellt, inwieweit die gestellten Anforderungen erfüllt werden.

Die Arbeit wird in Kapitel 7 mit einer Zusammenfassung und einem Ausblick abgeschlossen. Letzterer zeigt weitergehende Fragestellungen für zukünftige Forschung auf, die direkt oder indirekt mit der Prognose von Weichenstörungen zusammenhängen.

In Ergänzung zu den Themen der benannten Kapitel enthält der Anhang Formeln sowie zusätz-

liche Erläuterungen, Abbildungen und Tabellen. Sie dienen dem besseren Verständnis einzelner Aspekte. An den relevanten Stellen verweist der Text auf die jeweiligen Teile des Anhangs.

2 Zustandsprädiktion von Eisenbahnweichen im technischen und wirtschaftlichen Kontext

Weichen stehen mit dem Gesamtsystem Eisenbahn in einem bestimmten technischen, prozeduralen und ökonomischen Kontext. Auch die Zustandsprädiktion steht in diesem Kontext. Dieses Kapitel soll sowohl die Weichen in das Gesamtsystem einordnen, als auch deren Aufbau und die Funktion erläutern (Abschnitt 2.1). Dadurch wird die Thematik in einen breiteren Kontext gestellt, um so das grundlegende Verständnis zu erleichtern. Des Weiteren wird auf die Prädiktion als Mittel zur Erhöhung der Wirtschaftlichkeit im Rahmen der Instandhaltung eingegangen (Abschnitt 2.2), um die Relevanz dieser Arbeit zu verdeutlichen. Insbesondere wird dazu die zustandsorientierte Instandhaltung bei der Eisenbahninfrastruktur beleuchtet (Abschnitt 2.3), wodurch die Motivation für die vorliegende Arbeit dargelegt wird. Zum Abschluss des Kapitels werden bestehende Prädiktionsmodelle für Weichen analysiert (Abschnitt 2.4), um herauszufinden inwieweit diese bereits die Ziele dieser Arbeit erfüllen.

2.1 Infrastrukturelement Weiche

Zunächst sei festgehalten, dass die Begriffe Eisenbahn und Eisenbahninfrastruktur hier im Sinne der Definition nach §1 und §2 des Allgemeinen Eisenbahngesetzes (AEG) verwendet werden.

„(1) Eisenbahnen sind öffentliche Einrichtungen oder privatrechtlich organisierte Unternehmen, die Eisenbahnverkehrsleistungen erbringen oder eine Eisenbahninfrastruktur betreiben.

(2) Eisenbahnverkehrsleistungen sind die Beförderung von Personen oder Gütern auf einer Eisenbahninfrastruktur. Eisenbahnverkehrsunternehmen müssen in der Lage sein, die Zugförderung sicherzustellen. . . .

(3) Die Eisenbahninfrastruktur umfasst die Betriebsanlagen der Eisenbahnen einschließlich der Bahnstromfernleitungen.

(3a) Betreiber der Schienenwege ist jedes Eisenbahninfrastrukturunternehmen, das den Betrieb, den Bau und die Unterhaltung der Schienenwege der Eisenbahn zum Gegenstand hat, mit Ausnahme der Schienenwege in Serviceeinrichtungen.“ [Bun93, S. 1 ff.] Dabei gelten, gemäß §1 Abs. 2 AEG, nicht als Eisenbahnen im Sinne des Gesetzes „... andere Schienenbahnen wie Magnetschwebebahnen, Straßenbahnen und die nach ihrer Bau- oder Betriebsweise ähnlichen Bahnen, Bergbahnen und sonstige Bahnen besonderer Bauart.“ [Bun93, S. 1]

Im Zuge dieser Definition werden auch nur Weichen betrachtet, die Teil der Infrastruktur von Eisenbahnen sind, um den Untersuchungsraum auf häufige und ähnliche Objekte einzugrenzen. Anderen Bahnen sind durch andere Sicherheitsanforderungen und andere Konstruktionen charakterisiert, wodurch sich die mechanische Konstruktion, Stellkräfte und Antriebstechniken erheblich von Eisenbahnweichen unterscheiden können.

2.1.1 Einordnung in das Gesamtsystem Eisenbahn

Mit dem folgenden Überblick wird die Weiche in das Gesamtsystem Eisenbahn eingeordnet. Aussagen zu bestimmten Komponenten und deren Zustandsüberwachung, die in dieser Arbeit getätigt werden, lassen sich so besser nachvollziehen. Das Gesamtsystem Eisenbahn wird auf oberster Ebene oft in Bahnbetrieb, Personal, Rollmaterial und Infrastruktur unterteilt. Diese Gliederung folgt der Eisenbahn-Bau- und Betriebsordnung (EBO) [Bun12], auch wenn dort die Begriffe Fahrzeuge statt Rollmaterial und Bahnanlagen statt Infrastruktur verwendet werden. Weichen sind Elemente der Infrastruktur, weshalb die anderen Bestandteile hier nicht weiter detailliert werden. Soll die Infrastruktur weiter aufgegliedert werden, so orientieren sich Lehrbücher oft an der Richtlinie 836 „*Erdbauwerke und sonstige geotechnische Bauwerke planen, bauen und instand halten*“ der DB Netz AG [DB 99], die eine Unterteilung in Oberbau, Unterbau und Untergrund vornimmt. Des Weiteren behandeln einige Standardwerke wie bspw. das „*Handbuch Eisenbahninfrastruktur*“ [Fen07] oder „*Bahnwesen*“ [Fie99] Fahrleitungsanlagen, Telekommunikationsanlagen, Bahnübergänge (BÜ), Ingenieurbauwerke und die Eisenbahnleit- und -sicherungstechnik (LST) als eigenständige Kapitel, so dass diese Einteilung für eine weitere Untergliederung der Infrastruktur herangezogen wird. An dieser Stelle spaltet sich die Betrachtung von Weichen auf, da einige Weichenkomponenten dem Oberbau und einige der LST zugeordnet sind. Diese historisch gewachsene Aufteilung spiegelt sich auch in den jeweiligen Regelwerken zur Planung, zum Bau und zur Instandhaltung des Oberbaus und der LST wider.

Der Oberbau setzt sich aus Planumsschicht, Bettung, Schwellen, Schienenbefestigung und Schienen zusammen, oder besteht alternativ aus einer festen Fahrbahn, die mehrere dieser Elemente integriert [Fen07, S. 90]. Hinzu kommen noch Gleisabschlüsse und Gleisverbindungen. Zu letzteren zählen Schiebebühnen, Drehscheiben, Kreuzung und eben Weichen [Fie99, S. 123]. Weichen sind für den Eisenbahnbetrieb unverzichtbar. Sie verbinden Gleise, wodurch überhaupt erst ein Netz entsteht. Im Gegensatz zu Drehscheiben oder Schiebebühnen erlauben sie den Gleiswechsel ohne Fahrunterbrechung.

Die LST setzt sich aus Zugbeeinflussungsanlagen, Bahnübergangssicherungsanlagen (BÜSA) und Stellwerken zusammen [FNT11]. Dabei sind nur letztere im Zusammenhang mit Weichen relevant. Einfach ausgedrückt bestehen Stellwerke aus einer Innenanlage und einer Außenanlage. Die Innenanlage fasst vor allem die Technik zur Sicherungslogik und Fahrwegeinstellung zusammen, während die Außenanlage die physischen Elemente zur Umsetzung dieser Sicherungslogik und Fahrwegeinstellung aggregiert. Mit Ausnahme von ortsgestellten Weichen werden Weichen zwar über die Innenanlage bedient, für die vorliegende Arbeit ist sie jedoch nicht von Bedeutung, weil sie höchst selten Ursache einer Weichenstörung ist. Der Außenanlage werden die Gleisfreimeldung, ortsfeste Signale und die beweglichen Fahrwegelemente zugeordnet. Zu letzteren zählen das Weichenstellsystem sowie Gleissperren, bewegliche Brücken, Drehscheiben, Schiebebühnen und Tore. [Fen07, S. 610].

Das Weichenstellsystem wird also der LST zugeordnet, während die restlichen Weichenkomponenten dem Oberbau zugeordnet werden. Die beschriebene Einordnung in das Gesamtsystem Eisenbahn ist in Abbildung 2-1 noch einmal veranschaulicht. Daraus wird auch ersichtlich, wie vielfältig und komplex das System Eisenbahn ist. Da für die vorliegende Arbeit nur ein kleiner Teil relevant ist, wird auf weitere Ausführungen verzichtet. Für detaillierte Erläuterungen oder

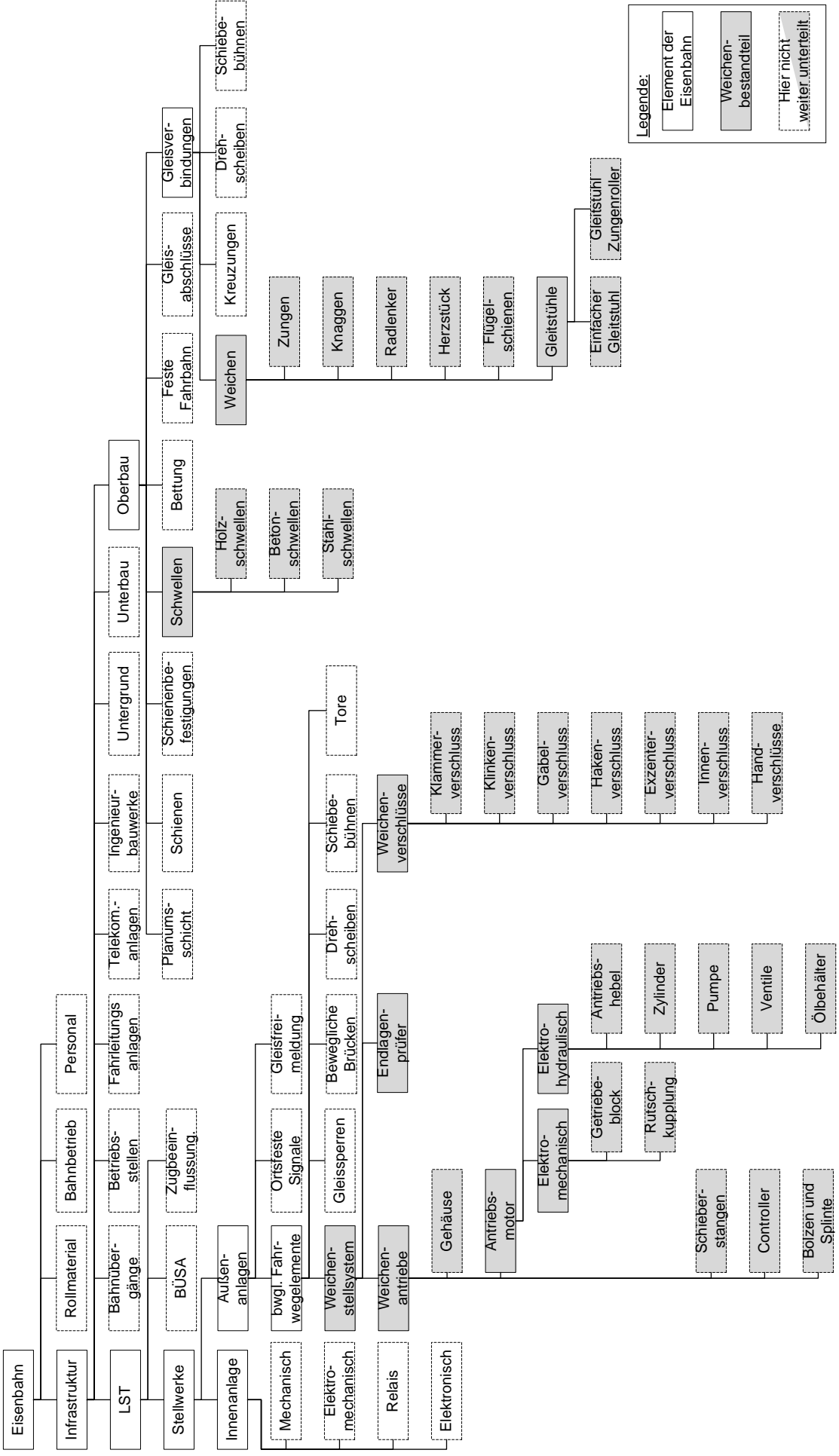


Abbildung 2-1: Übersicht zur Einordnung von Weichenkomponenten in die Struktur des Systems Eisenbahn

genaue Definitionen der Eisenbahninfrastrukturelemente sei auf die umfassenden Grundlagenwerke von Fiedler [Fie99], Naumann und Pachl [NP04], Fendrich [Fen07], Pachl [Pac09] und Fenner et al. [FNT11] verwiesen.

2.1.2 Aufbau und Funktion von Eisenbahnweichen

Nach dem Überblick zum Gesamtsystem erläutert dieser Abschnitt den Aufbau und die Funktion von Weichen. Es wird detailliert auf die einzelnen Komponenten eingegangen und dann der Weichenstellvorgang erläutert, um den Leser mit den relevanten Fachbegriffen und Zusammenhängen vertraut zu machen.

Weichen treten in unterschiedlichen Typen auf [Fie99, S. 124]. Die folgenden Weichentypen sind in Abbildung 2-2 schematisch dargestellt:

- Die **einfache Weiche (EW)** besteht aus einem geraden Gleis der Hauptrichtung und einem bogenförmig abzweigenden Gleis in linker oder rechter Richtung.
- Die **Doppelweiche (DW)** besteht aus ineinander verschobenen Weichen. DW gibt es in einseitiger Ausfertigung mit beiden Abzweiggleisen in einer Richtung oder als zweiseitige Ausfertigung mit Abzweiggleisen in unterschiedliche Richtungen.
- Die **Innenbogenweiche (IBW)** ist eine Weiche, bei der sowohl das Hauptgleis als auch das Abzweiggleis bogenförmig verlaufen. Beide Mittelpunkte der Bögen liegen auf der gleichen Seite der Strecke.
- Bei der **Außenbogenweiche (ABW)** befinden sich beide Mittelpunkte der Bögen auf verschiedenen Seiten.
- Die **einfache Kreuzungsweiche (EKW)** ist die Verbindung von einer Kreuzung und einer Weiche. Ein Gleiswechsel ist nur von einem Teil aus möglich.
- Die **doppelte Kreuzungsweiche (DKW)** ist hingegen die Verbindung einer Kreuzung mit zwei Weichen, so dass ein Gleiswechsel von jedem Teil aus möglich ist.

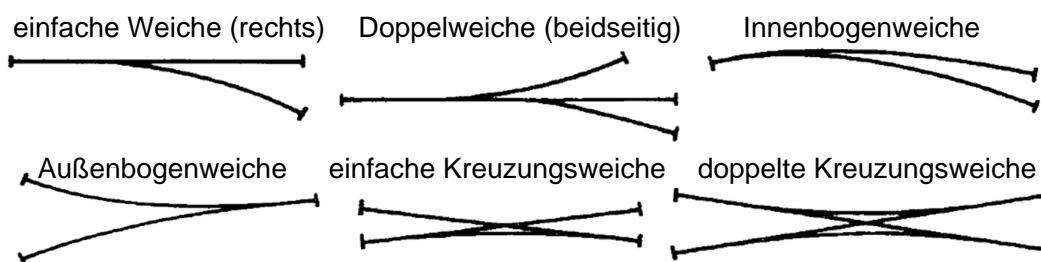


Abbildung 2-2: Schematische Darstellung der Weichentypen (in Anlehnung an [Fie99, S. 124])

Weil DW, EKW und DKW aufwendiger bei der Instandhaltung und der sicherungstechnischen Einbindung sind, werden diese Typen tendenziell nur dann verbaut, wenn die begrenzten räumlichen Verhältnisse der Strecke den Einsatz notwendig werden lassen [Hom10, S. 48].

Am Beispiel der EW lassen sich die einzelnen Konstruktionsbestandteile (siehe Abbildung 2-3) und ihre Funktion gut erläutern. Aber auch bei anderen Weichentypen sind diese Bestandteile zu finden, eventuell in leicht abweichender Form bzw. Anzahl.

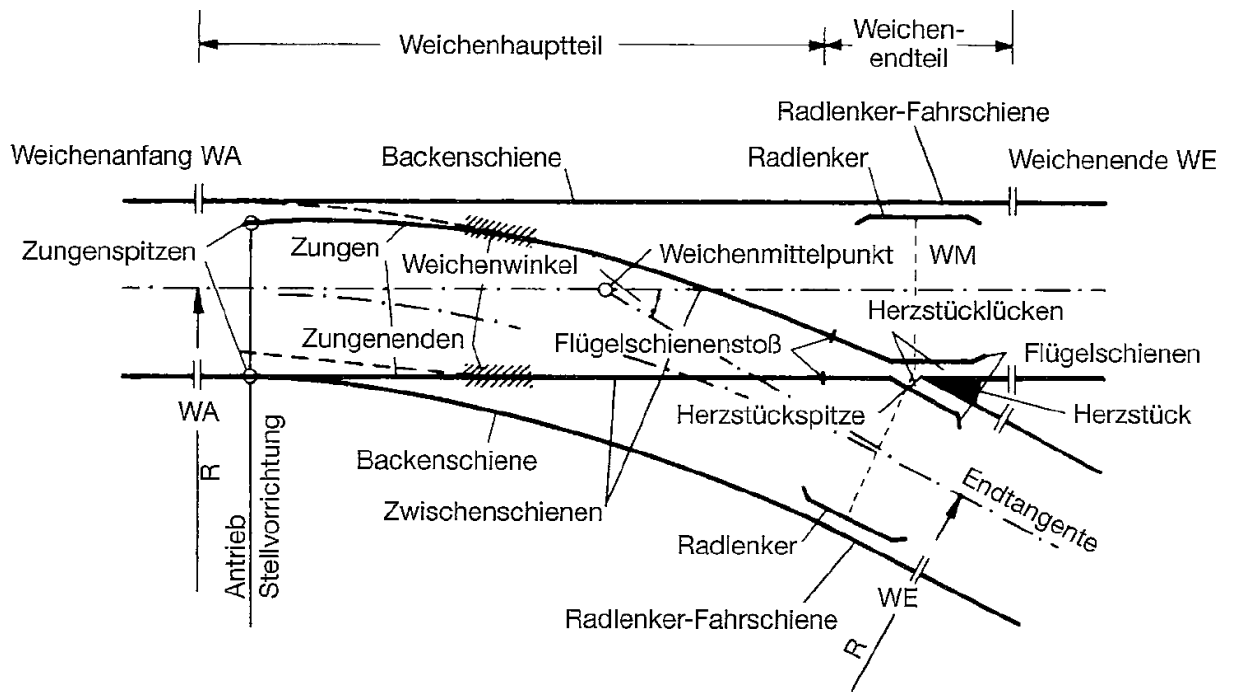


Abbildung 2-3: Konstruktionselemente einer einfachen Weiche in schematischer Darstellung (Copyright: Werner Verlag in [FS12, S. 127])

Die **Backenschienen** befinden sich außen an Haupt- und Abzweiggleis zur Spurführung des aufliegenden Rades und sind an der festen Fahrbahn oder den Schwellen befestigt. Letztere können aus Holz, Beton oder Stahl gefertigt sein.

Die **Zungen** sind die beweglichen Bestandteile an der Weichenspitze. Sie werden je nach Richtungswunsch an die eine oder andere Backenschiene angelegt. Die Zunge, welche zum abzweigenden Gleis gehört, läuft bogenförmig auf die andere Zunge zu. Beide bilden so einen Drehpunkt, um den die Zungen geschwenkt werden. Dabei werden entweder Gelenkzungen mit einem tatsächlichen Gelenk verwendet oder Federzungen, bei denen die Elastizität des Werkstoffs ausgenutzt wird [Jan09, S. 51]. Da die Zungen bewegt werden müssen, können sie nicht an den Schwellen oder der festen Fahrbahn fixiert werden. Um dennoch eine gewisse Stabilität gegen ungewollte Bewegung oder Verformung durch die starken lateralen Kräfte bei der Zugüberfahrt zu geben und einen ausreichenden Abstand (mindestens 10 mm) zwischen abliegender Zunge und Backenschiene zu garantieren, werden Knaggen als Abstandshalter an die Backenschiene montiert [Fie99, S. 132].

Im **Herzstück** laufen die Schienenstränge von Haupt- und Abzweiggleis zusammen. Damit der Spurkranz des Fahrzeugradreifens an dieser Stelle passieren kann, muss die Schiene zwischen den Zungen und dem Herzstück unterbrochen sein (die sogenannten Herzstückklücken). Weil durch die Fahrkanten- und Fahrflächenunterbrechung die Spurführung nicht mehr vollständig gegeben ist, übernehmen die **Radlenker** an den Backenschienen und die Flügelschienen diese Funktion. Wenn bewegliche Herzstückspitzen verbaut werden, schließen diese die Herzstückklücken, so dass eine unterbrechungsfreie Fahrkante gegeben ist.

Die Zungen selbst liegen auf **Gleitstühlen** (vgl. Abbildung 2-4) auf. In der Regel gleitet der Schienenfuß auf einer Metallplatte. Zur Verringerung des Reibungswiderstands muss diese re-

gelmäßig gereinigt und geschmiert werden. Zur Einsparung von Schmierstoffen und zur Verringerung des Wartungsaufwands werden deshalb immer häufiger Gleitstühle mit **Zungenrollern** verbaut. Abbildung 2-4 zeigt ein Beispiel für einen einfachen Gleitstuhl (a) und einen Gleitstuhl mit Zungenrollern (b).

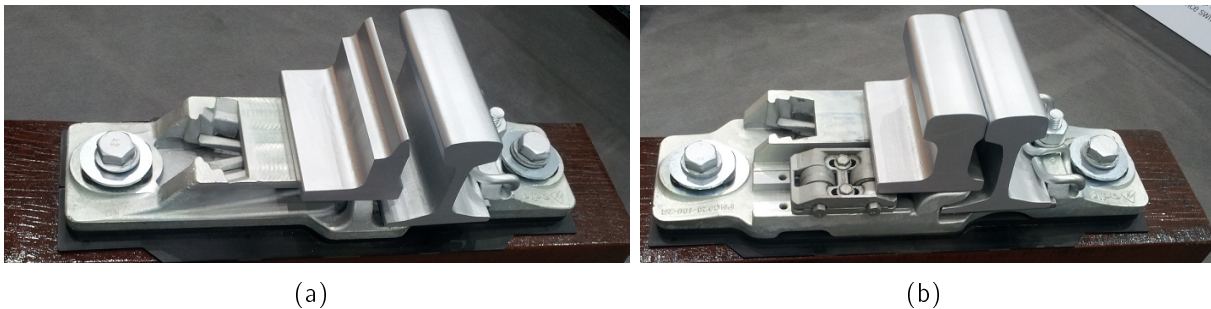


Abbildung 2-4: Beispiel eines einfachen Gleitstuhls (a) und eines Gleitstuhls mit Zungenrollern (b)

Der **Antrieb** der Weiche bewegt mittels Schieberstangen die Zungen in ihre jeweilige Stellung. Zu finden sind ortsgestellte Weichen, die per Hand oder per elektrischem Antrieb umgestellt werden. Bei ferngesteuerten Weichen wird der Antrieb aus einem Stellwerk heraus bedient. Es gibt zwar innenliegende, also im Gleis liegende Antriebe, aber in Deutschland ist der Antrieb in der Regel neben dem Gleis in einem feuerverzinkten Graugussgehäuse untergebracht [Jan09, S. 52] (siehe dazu Abbildung 2-5).

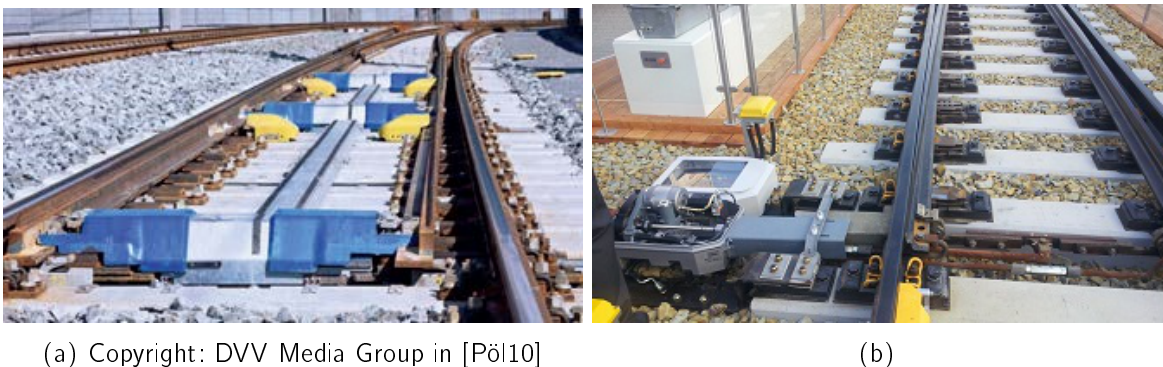


Abbildung 2-5: Beispiel für einen innenliegenden (a) und einen neben dem Gleis liegenden Antrieb (b)

Ein Elektromotor treibt die Schieberstangen entweder auf mechanischem Wege an oder über ein Pumpensystem für eine hydraulische Kraftübertragung. Im Fall des elektromechanischen Antriebs schont eine Rutschkupplung den Motor vor Schäden, die bspw. durch einen blockierenden Fremdkörper zwischen Zunge und Backenschiene, beim plötzlichen Ende durch Erreichen der Endlage oder beim Auffahren der Weiche entstehen könnten. (Als Auffahren wird die Überfahrt vom Herzstück zur Zungenspitze bezeichnet, die aus dem nicht eingestellten Abzweig heraus geschieht, z.B. aus dem unteren Abzweig in Abbildung 2-3. Dabei drückt sich der Spurkranz zwischen anliegende Zunge und Backenschiene, so dass die momentane Endlage aufgekeilt wird.) Die Umwandlung der Drehbewegung des Motors in eine Längsbewegung der Schieberstangen erfolgt über einen Getriebeblock. Im Fall des elektrohydraulischen Antriebs erzeugt der Motor bzw. das angeschlossene Pumpensystem in zwei Zylindern jeweils einen Unter- und einen Überdruck, je nach Drehrichtung. Dadurch werden die Kolben bewegt, die wiederum die Schieberstangen bewegen. In ähnlicher Weise wie die Rutschkupplung begrenzt ein Ventilsystem die Stellkraft

und schützt den Antrieb vor Schäden durch Fremdkörper oder beim Auffahren. [FNT11, S. 92] Abbildung 2-6 zeigt den elektromechanischen Antrieb S700K der Firma Siemens (a) und den elektrohydraulischen Antrieb L826H der Firma Alcatel (heute Thales) (b). Bei beiden Antriebsarten sind elektronische Controller zur Steuerung im Gehäuse untergebracht.

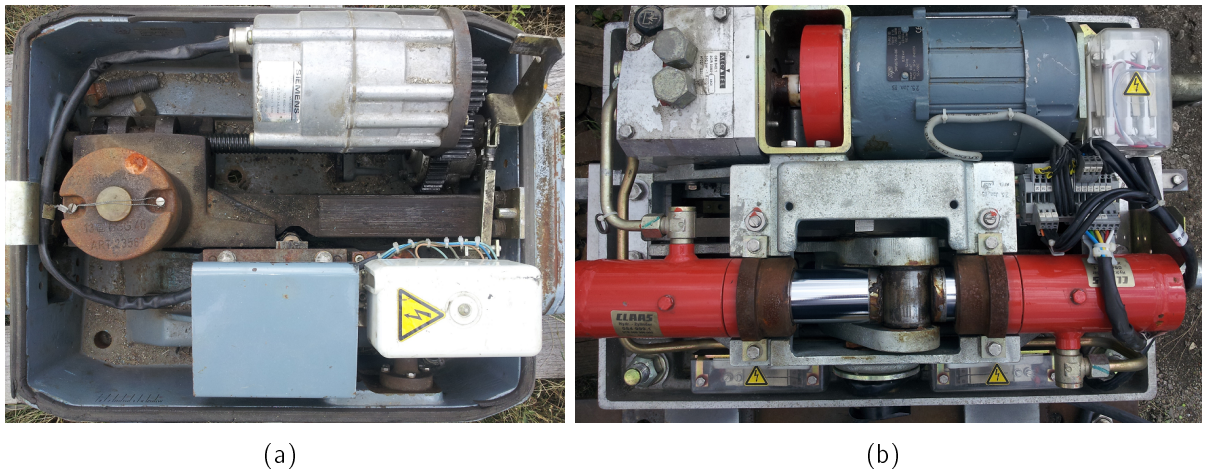


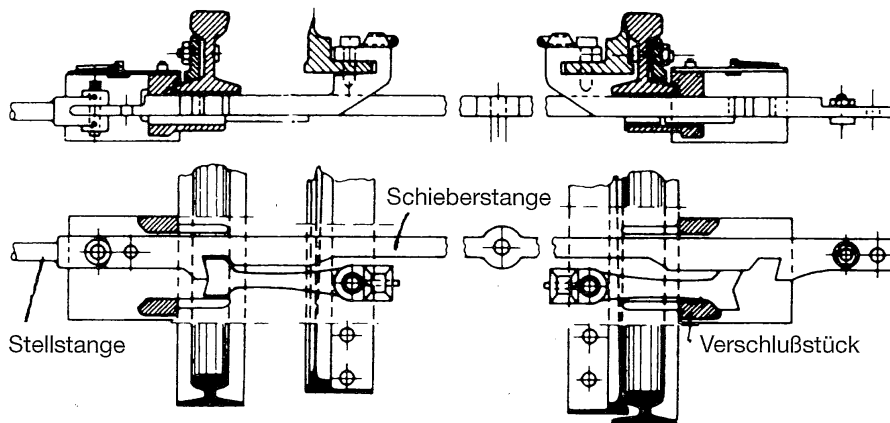
Abbildung 2-6: Elektromechanischer Weichenantrieb S700K von Siemens (a) und elektrohydraulischer Antrieb L826H von Alcatel (b)

In Deutschland und insbesondere bei der DB Netz AG werden elektrisch angetriebenen Weichen über eine Vierdraht-Schaltung mit Energie versorgt. Dabei werden drei Adern für die eigentliche Versorgung des Motors genutzt (Drehstrommotor mit $3 \times 400 \text{ V}$) und eine Ader für die Steuerungsinformation [FNT11, S. 91, S. 99].

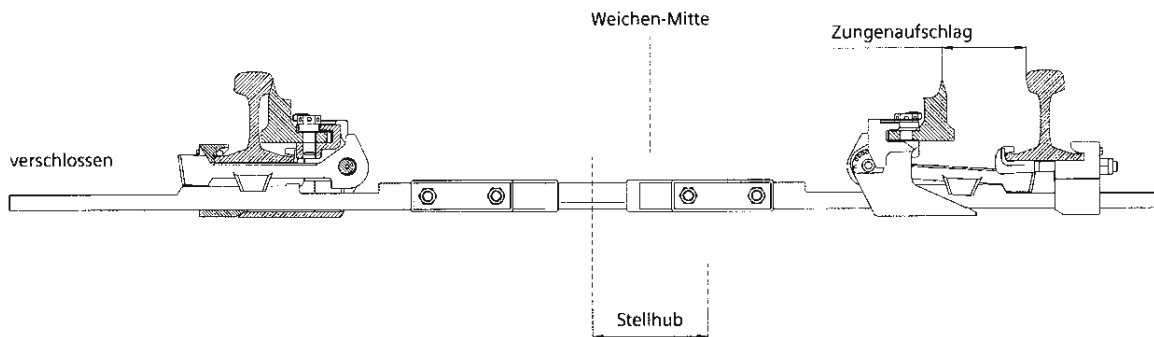
Die Erfüllung hoher Sicherheitsanforderungen macht die Bahn zwar zu einem der sichersten Verkehrsmittel, stellt aber auch besondere Anforderungen an die Technik. Bei Weichen ist dies die sichere Überwachung der Zungenposition in Endlage, d.h. eine an der Backenschiene anliegende Zunge (Abstand kleiner 3 mm) und eine abliegende Zunge (Abstand mindestens 10 mm). Dadurch wird verhindert, dass ein Zug bei der Überfahrt entgleist. Aus demselben Grund muss die Weiche auch gegen Umstellen unter dem fahrenden Zug in der Endlage festgehalten werden [Bun12, § 14 Abs. 9]. Die Art der Endlagenprüfung ist abhängig von der Stellwerksform, an die ein Weichenstellsystem angeschlossen ist. So wird bei rein mechanisch angetriebenen Weichen, wie ortsgestellten Weichen oder bei mechanischen Stellwerken, ein Weichenriegel oder ein Zungenprüfer eingesetzt. Auf mechanischem Wege verhindern diese das vollständige Umlegen des entsprechenden Hebels im Stellwerk, wodurch auch der Fahrtbegriff nicht eingestellt werden kann [Hom10, S. 50 f.]. Bei den weitaus häufigeren elektrisch angetriebenen Weichen wird die Endlage durch Schaltkontakte überprüft, dem sogenannten **Endlagenprüfer (ELP)**. Zum Schutz der Kontakte sind diese in einem eigenen oder im Gehäuse des Antriebs untergebracht. [FNT11, S. 95]

Wie bereits erwähnt, besteht bei Weichenüberfahrten ein erhöhtes Entgleisungsrisiko, zum einen durch die ohnehin vom Sinuslauf der Räder induzierten lateralen Kräfte auf die Schiene bzw. Zunge, zum anderen durch die hinzukommenden lateralen Kräfte bei der Richtungsänderung zum Eintritt in die oder Austritt aus der Weichenkurve (für den Passagier als deutlicher Ruck wahrnehmbar). In Verbindung mit den vertikalen Kräften durch die Zugmasse kann es vorkommen, dass die Zunge angehoben und gesenkt wird, bis zu einem Punkt, an dem der Abstand zwischen

Zunge und Backenschiene keine Spurführung mehr gewährleistet. Der **Weichenverschluss** verhindert, dass es zu einer solchen Situation kommt. Er befindet sich an der Zungenspitze.



(a) Copyright: Werner Verlag in [Fie99, S. 136]



(b) In Anlehnung an [Sie12]

Abbildung 2-7: Zeichnung eines Klammerverschlusses (a) und des Klinkenverschlusses CKA der Siemens Schweiz AG (b)

Seit den frühen Jahren der Eisenbahn wurden zahlreiche Verschlussarten in unterschiedlichen Bauformen entwickelt. Viele davon wurden mit dem technischen Fortschritt durch verbesserte Konstruktionen abgelöst. Im Folgenden sollen die häufig im Netz der Deutsche Bahn Aktiengesellschaft (DB AG) zu findenden Verschlussarten aufgezählt werden:

- Der am häufigsten in Deutschland verbaute Verschluss an der Zungenspitze ist der **Klammerverschluss** [Fen07, S. 611]. Er ist in Abbildung 2-7(a) dargestellt. An jeder Seite läuft dabei eine schwalbenschwanzähnliche Klammer, welche mit der Zunge verbunden ist, durch ein Verschlussstück hindurch. Die Klammer ragt bei der anliegenden Zunge aus dem Verschlussstück heraus und drückt sowohl gegen die an dieser Stelle verbreiterte Schieberstange als auch gegen die Außenseite des Verschlussstücks. Die Zunge ist dadurch mechanisch gegen eine Bewegung verschlossen. Bei der abliegenden Zunge ist die Klammer in die dafür vorgesehene Aussparung an der Schieberstange eingerastet, so dass beide im Verschlussstück unter der Schiene Platz finden. Die Klammer kann im Verschlussstück nicht aus der Aussparung heraustreten. So festgehalten ist der Abstand zwischen Zunge und Schiene fixiert. Durch die Konstruktion des Verschlussstücks und der Klammer kann es bei Bewegungen der Zunge in Fahrtrichtung (Zungenwanderung) zu Störungen durch ungewollte Verklemmung kommen. Der Klammerverschluss kann nicht verhindern, dass sich die Zungenposition vertikal verändert. Außerdem ist der Klammerverschluss anfällig für

Störungen durch vereiste oder an den Schienen festgefrorene Zungen. [FNT11, S. 106 f.]

- Etwas jünger ist der **Klinkenverschluss**, welcher vom Prinzip genauso funktioniert wie der Klammerverschluss. Abbildung 2-7(b) zeigt diese Verschlussart. Durch die Drehung der Verschlusseinrichtung um 90° wird jedoch die vertikale Bewegung der Zungen verhindert. Dadurch ist der Verschluss besser für Weichen geeignet, die mit hohen Geschwindigkeiten überfahren werden. [FNT11, S. 107] Auch der Klinkenverschluss ist anfällig für Störungen durch Zungenwanderung sowie durch Schnee und Eis.
- Der **Gabelverschluss** wird ausschließlich als zusätzlicher Verschluss bei Weichen eingesetzt, die wegen ihrer Länge an mehreren Stellen verschlossen werden müssen. Gabelverschlüsse können nicht als Spitzenverschlüsse eingesetzt werden, da sie die Weiche nicht festhalten, sondern lediglich den definierten Abstand zwischen Zunge und Schiene gewährleisten. Im Allgemeinen wird der Gabelverschluss durch den Antrieb an dem eigentlichen Verschluss mitbewegt. Er ist kaum anfällig für Störungen durch Zungenwanderung oder Schnee und Eis. [FNT11, S. 108]
- **Handverschlüsse** kommen immer dann zum Einsatz, wenn Weichen vorübergehend vor Ort gegen das Umstellen gesichert werden müssen [Hom10, S. 51], z.B. bei Störungen oder Bauarbeiten. Sie werden an Schiene und Zunge direkt befestigt und mit einem Schlüssel von Hand verschlossen, so dass die Weiche in dieser einen Position gehalten wird und auch nicht mehr auffahrbar ist. Da Handverschlüsse nur absichtlich und geplant eingesetzt werden, spielen sie für die Zustandsprädiktion allerdings keine Rolle.

Alle genannten Verschlussarten sind prinzipiell bei allen Weichen einsetzbar. Die Verschlussarten und Bauformen der Hersteller unterscheiden sich jedoch bezüglich ihrer Fähigkeit, den lateralen Kräften entgegenzuwirken. Außerdem unterscheiden sie sich in ihrer Toleranz gegenüber Schnee oder Eis und gegenüber Materialausdehnung auf Grund von Temperaturveränderungen, genannt Zungenwanderung. [FNT11, S. 105]

Je höher die Fahrgeschwindigkeiten, desto höher sind die Kräfte, die bei einer Zugfahrt über eine Weiche wirken. Damit die Festhaltekraft der Verschlüsse das Entgleisungsrisiko ausreichend klein hält, dürfen Weichen nur mit einer bestimmten Maximalgeschwindigkeit überfahren werden, die von ihrer Länge und demzufolge ihrem Radius (oder dem analogen Abzweigwinkel) abhängt. Aber auch auf Grund des Fahrkomforts ist die Geschwindigkeit begrenzt, damit die seitliche Beschleunigung – der Ruck – für Reisende nicht zu groß ist [Zwa09, S. 10]. Die längeren Weichen benötigen mehrere Antriebe und Verschlüsse, um die längeren und demzufolge schwereren Zungen zu bewegen und in der Endlage festzuhalten. Dabei kommen mitunter auch separate ELP vor, die neben dem Gleis in eigenen Gehäusen untergebracht sind.

2.1.3 Umstellvorgang einer Eisenbahnweiche

Im vorherigen Abschnitt wurden der Aufbau von Weichen und die Komponenten mit ihrer jeweiligen Funktion näher erläutert. Nun, da die Fachbegriffe eingeführt wurden, kann auch der Umstellvorgang eindeutig beschrieben werden. Im Allgemeinen wird der Umstellvorgang einer Weiche in drei Phasen unterteilt. Bei einem Klammer- oder Klinkenverschluss laufen diese folgendermaßen ab (siehe auch [Fen07, S. 612]):

1. In der **Entriegelung** wird die abliegende Zunge von der Schieberstange bewegt, während der Verschluss der anliegenden Zunge geöffnet wird.
2. In der **Bewegung beider Zungen**, auch als Stellung bezeichnet, ist keine Klammer in einer Verschlussposition. Beide Zungen werden vom Antrieb gleichzeitig über die Gleitstühle bewegt.
3. Es folgt die **Verriegelung**, bei der die anfangs abliegende Zunge bereits an der Schiene anliegt und sich nur noch die dazugehörige Verschlussklammer und die jetzt neu abliegende Zunge bewegen. Der Vorgang endet, wenn schließlich die Klammer der jetzt anliegenden Zunge ihre Verschlussposition erreicht und auch die Klammer der jetzt abliegenden Zunge in die entsprechende Aussparung der Schieberstange eingerastet ist.

Wie Abbildung 2-8 verdeutlicht, wird dabei ein Stellweg von 220 mm überbrückt. (Noch anschaulicher wird der Umstellvorgang in der Animation in Lindtner et al. [LRLT09, Kapitel 321] dargestellt.) Für den Umstellvorgang benötigen die gängigen Antriebe vier bis sechs Sekunden. Es sei erwähnt, dass sich die Phasen noch weiter unterteilen lassen, wenn der Leistungsverlauf des Motors betrachtet wird, was allerdings erst im späteren Verlauf der Arbeit wichtig wird.

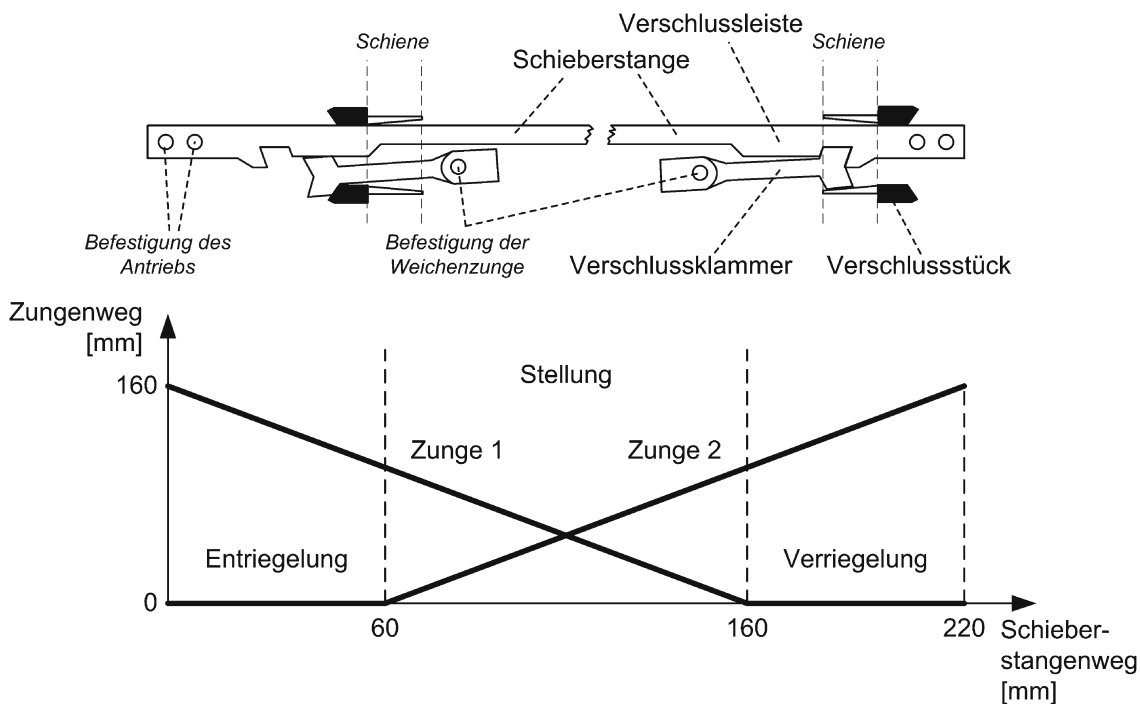


Abbildung 2-8: Umstellvorgang einer einfachen Weiche mit Klammerverschluss (Copyright: Springer Verlag in [Fen07, S. 611])

2.2 Instandhaltung

Nachdem der vorherige Abschnitt den Weichenaufbau und die Funktion erläutert und in einen technischen Zusammenhang mit dem System Eisenbahn gestellt hat, geht der folgende Abschnitt auf den wirtschaftlichen Kontext ein. Dadurch wird ersichtlich, wie die Lebenszykluskosten (insbesondere die Instandhaltungskosten), die Instandhaltungsstrategie und die Zustandsprognose miteinander verknüpft sind. Erst in Verbindung mit den erläuterten Fachbegriffen wird es möglich, zu definieren, wie eine Zustandsprognose in dieser Arbeit verstanden wird.

2.2.1 Abnutzungsvorrat, verbleibende Nutzungsdauer und Instandhaltungsmaßnahmen

Bevor anhand der Instandhaltungsstrategien der Zusammenhang zwischen Instandhaltungskosten und Zustandsprognose hergestellt werden kann, werden die Begriffe des Abnutzungsvorrats, der verbleibenden Nutzungsdauer und der Instandhaltungsmaßnahmen benötigt. In dieser Arbeit werden die sehr treffenden Definitionen der DIN 31051 verwendet. Sie definiert den Abnutzungsvorrat als einen:

„Vorrat der möglichen Funktionserfüllung unter festgelegten Bedingungen, der einer Betrachtungseinheit aufgrund der Herstellung, Instandsetzung oder Verbesserung innewohnt“ [DIN06, S. 5].

Als Synonym wird der Begriff der verbleibenden Nutzungsdauer verwendet, der vor allem im Englischen als Remaining Useful Life (RUL) gebräuchlich ist und explizit den Zeitbezug ausdrückt. Als eher theoretische Größe besitzt der Abnutzungsvorrat nach der Herstellung ein Maximum und zum Zeitpunkt des Ausfalls den Wert Null. Der Verlauf des Abnutzungsvorrates, also die Abnutzung, zwischen Herstellung und Ausfall kann sehr unterschiedlich ausfallen, u.a. als monoton fallende, lineare Funktion, als regressive Funktion, als progressive Funktion, etc. Wie der Abnutzungsvorrat verläuft, bestimmt sich aus den Betriebsbedingungen und der Instandhaltung.

Die Instandhaltung ist definiert als:

„Kombination aller technischen und administrativen Maßnahmen sowie Maßnahmen des Managements während des Lebenszyklus einer Betrachtungseinheit zur Erhaltung des funktionsfähigen Zustandes oder der Rückführung in diesen, so dass sie die geforderte Funktion erfüllen kann“ [DIN06, S. 3].

Die folgenden Begriffe sind als Maßnahmen der Instandhaltung näher beschrieben und beeinflussen den Abnutzungsvorrat:

- Als **Wartung** werden die Aktivitäten bezeichnet, die eine Verringerung des Abnutzungsvorrates verlangsamen.
- Aktivitäten, die den Istzustand einer Betrachtungseinheit ermitteln oder beurteilen, werden als **Inspektion** bezeichnet.
- Wird eine Betrachtungseinheit zurück in ein funktionsfähigen Zustand versetzt, wird dies als **Instandsetzung** verstanden.
- Unter **Verbesserung** werden alle Aktivitäten zusammengefasst, die die Leistungsfähigkeit einer Betrachtungseinheit erhöhen, ohne dass die eigentliche Funktion verändert wird.

2.2.2 Lebenszyklus und Lebenszykluskosten

Wie fast die gesamte Eisenbahninfrastruktur gehören Weichen zu den Investitionsgütern, die sich durch

- hohe Sicherheitsanforderungen,

- hohe Verfügbarkeitsanforderungen,
- hohe Anschaffungskosten und
- eine lange Lebensdauer

auszeichnen. Sie haben damit einen starken Einfluss auf die Wirtschaftlichkeit des Eisenbahninfrastrukturunternehmens (EIU) und im weiteren Sinne auch auf die Wirtschaftlichkeit des Schienenverkehrs. Die Investitions- und Instandhaltungsentscheidungen sind aber nicht nur aus betriebswirtschaftlicher Sicht relevant, sondern können im Fall der Infrastruktur auch zu erheblichen volkswirtschaftlichen Schäden führen, die aus der Nichtverfügbarkeit folgen [Ada89, S. 56].

Quasi standardmäßig wird für die ökonomische Betrachtung solcher Investitionsgüter heute die Bewertung der Lebenszykluskosten (engl. Life Cycle Costs) (LCC) herangezogen, die laut DIN EN 60300-3-3 als die kumulierten Kosten eines Produkts über seinen Lebenszyklus definiert sind [DIN03, S. 6]. Unter Lebenszyklus ist dabei der Zeitraum vom Konzept bis zur Entsorgung einer Anlage zu verstehen (siehe Abbildung 2-9). Die Norm beschreibt detailliert die einzelnen Aggregationsstufen der LCC und den Aufbau eines LCC-Modells. Da für die vorliegende Arbeit nur ein kleiner Teil daraus relevant ist, nämlich die Instandhaltungskosten (siehe Abschnitt 2.2.3), wird zur vertiefenden Lektüre eben diese Norm [DIN03], die ausführliche Fachliteratur von Wübbenhorst [Wüb84] oder die sehr übersichtliche Aufstellung von Gutsche [Gut10, S. 19] empfohlen.

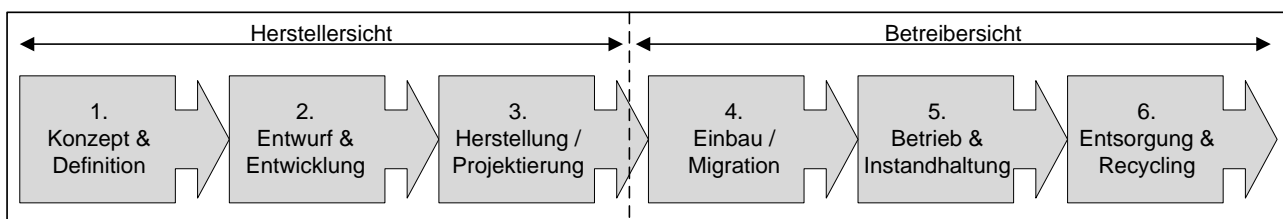


Abbildung 2-9: Phasen des Lebenszyklus (in Anlehnung an [DIN03, S. 8])

Nur Anlagen mit minimalen LCC können die Wettbewerbsfähigkeit eines Unternehmens auf Dauer sicherstellen [Wüb84]. In jeder Phase kann die Wirtschaftlichkeit der Anlagen mit unterschiedlichen Strategien oder mit entsprechenden Interventionen verbessert werden. Zum Beispiel werden bahntechnische Lösungsansätze auf Herstellerseite in Pölinger [Pöl10] präsentiert. Auf Betreiberseite werden Optimierungen für die Investition in Böhm und Scheier [BS10] oder für die Migration in Lackhove [Lac14] erläutert. Die Dissertation von Gutsche [Gut10] beschreibt eine integrierte Bewertung für die Investitions- und Instandhaltungsstrategie für die LST, während die Dissertation von Quiroga [Qui11] eine ganzheitliche Optimierung des Instandhaltungsprozesses der Gleisgeometrie bietet. In Lackhove et al. [LBS⁺11] sind aktuelle Entwicklungen bei der LST mit Blick auf die Betreiberseite zusammengefasst.

Allerdings sieht sich ein EIU in Bezug auf die LCC mit einem Dilemma konfrontiert. In der Entwicklungsphase einer Anlage werden deren Eigenschaften wie etwa die Montagefähigkeit, der Verschleiß oder die Instandhaltbarkeit determiniert. Es sind diese Eigenschaften, die dem Betreiber die Kosten gewissermaßen diktieren. D.h. in den frühen Phasen werden die LCC festgelegt, obwohl sie zum Großteil erst in den späten Phasen anfallen, wo die Möglichkeit zur Einflussnahme bereits stark abgenommen hat [Sch05, S. 44]. Zusätzlich gilt: Je länger die Lebensdauer ist, desto stärker bindet sich das EIU an die Kosten im Besitz. So ist es angesichts der langen Lebensdauer von Weichen nicht verwunderlich, dass die Ergebnisse des Projektes InfraCost deut-

lich gezeigt haben, dass für die partizipierenden EIU die Kosten von Betrieb und Instandhaltung bei Weichen der wesentliche Treiber für die LCC sind [BR02]. In ihrem Infrastrukturzustands- und -entwicklungsbericht für 2015 weist die DB Netz AG das Durchschnittsalter von Weichen mit 19.5 Jahren aus [DB 16]. Daraus folgt, dass die Möglichkeiten zur Kostensenkung für die Mehrheit der Weichen auf die Betriebsphase beschränkt ist und sich daran auch auf lange Sicht nichts ändern wird, noch dazu wo das Durchschnittsalter seit 2008 (19.5 Jahre) nicht reduziert wurde. Wie die Instandhaltungskosten definiert sind und welche Rolle die Instandhaltungsstrategie für diese Kosten spielt, wird in den nächsten Abschnitten beschrieben.

2.2.3 Instandhaltungskosten

Wie im vorherigen Abschnitt erläutert wurde, sind die Kosten der Instandhaltung Teil der Besitzkosten, die ihrerseits Teil der LCC sind. Die Instandhaltungskosten sind für eine Bewertung einer Zustandsprognose aus ökonomischer Sicht unverzichtbar, ein Zusammenhang, der in Kapitel 4.2 intensiv argumentiert wird. Hier soll zunächst ihre allgemeine Zusammensetzung beschrieben werden. Es werden direkte und indirekte Instandhaltungskosten unterschieden [Ada89]. Kosten, welche direkt bei der Durchführung einer Instandhaltungsmaßnahme anfallen, sind:

- die **Materialkosten** (inklusive Werkzeuge, Ersatzteile, etc.) und
- die **Personalkosten**, zusammengesetzt aus Personalbedarf, Personalkostensatz und Instandhaltungsdauer (inklusive Vor- und Nachbereitungszeiten).

Indirekte Kosten entstehen aus der Nichtverfügbarkeit als Folge der durchzuführenden Instandhaltungsmaßnahme oder des Ausfalls. Im Eisenbahnsektor hat sich der Begriff der Betriebser-schwerernskosten (BEK) als Synonym für die indirekten Kosten etabliert. Diese setzen sich in Anlehnung an Veit und Petri [VP08] zusammen aus:

- den **Opportunitätskosten** durch entgangene Einnahmen bzw. Gewinne,
- den Kosten als Folge von **Pönalen**,
- den Kosten aus **verpassten Anschlusszügen**,
- den Kosten für **Ersatzverkehr**,
- den zusätzlichen **Betriebsführungskosten**,
- den Kosten aus **negativen Marktreaktionen** und
- den **Umleitungskosten**.

2.2.4 Instandhaltungsstrategien

Bisher wurde erläutert, dass für ein Investitionsgut wie die Weiche die Instandhaltungskosten (Abschnitt 2.2.3) maßgeblicher LCC-Treiber (Abschnitt 2.2.2) sind. Die verschiedenen Instandhaltungsmaßnahmen verändern den Abnutzungsvorrat bzw. die RUL (Abschnitt 2.2.1). Als Instandhaltungsstrategie wird die grundlegende Art und Weise der Instandhaltung einer, mehrerer oder aller Betrachtungseinheiten bezeichnet. Sie entscheidet allgemein, bei welcher RUL welche Maßnahmen vollzogen werden und bestimmt demzufolge die direkten und indirekten Instandhaltungskosten. Die DIN EN 13306 definiert im wesentlichen zwei Arten von Strategien [DIN12].

Die **korrektive Instandhaltung** wird ausgeführt, wenn der Abnutzungsvorrat verbraucht ist, also das Schadensereignis (Fehler oder Ausfall) eingetreten und bekannt ist (siehe Abbildung 2-10 (a)). Es wird meist mit Instandsetzung reagiert. Während der Nutzung findet keinerlei Überwachung, Inspektion oder Wartung statt. Im wirtschaftlichen Sinne ist diese Strategie sinnvoll, wenn die Instandhaltungskosten gering sind und aus Schäden kein Sicherheitsrisiko entsteht.

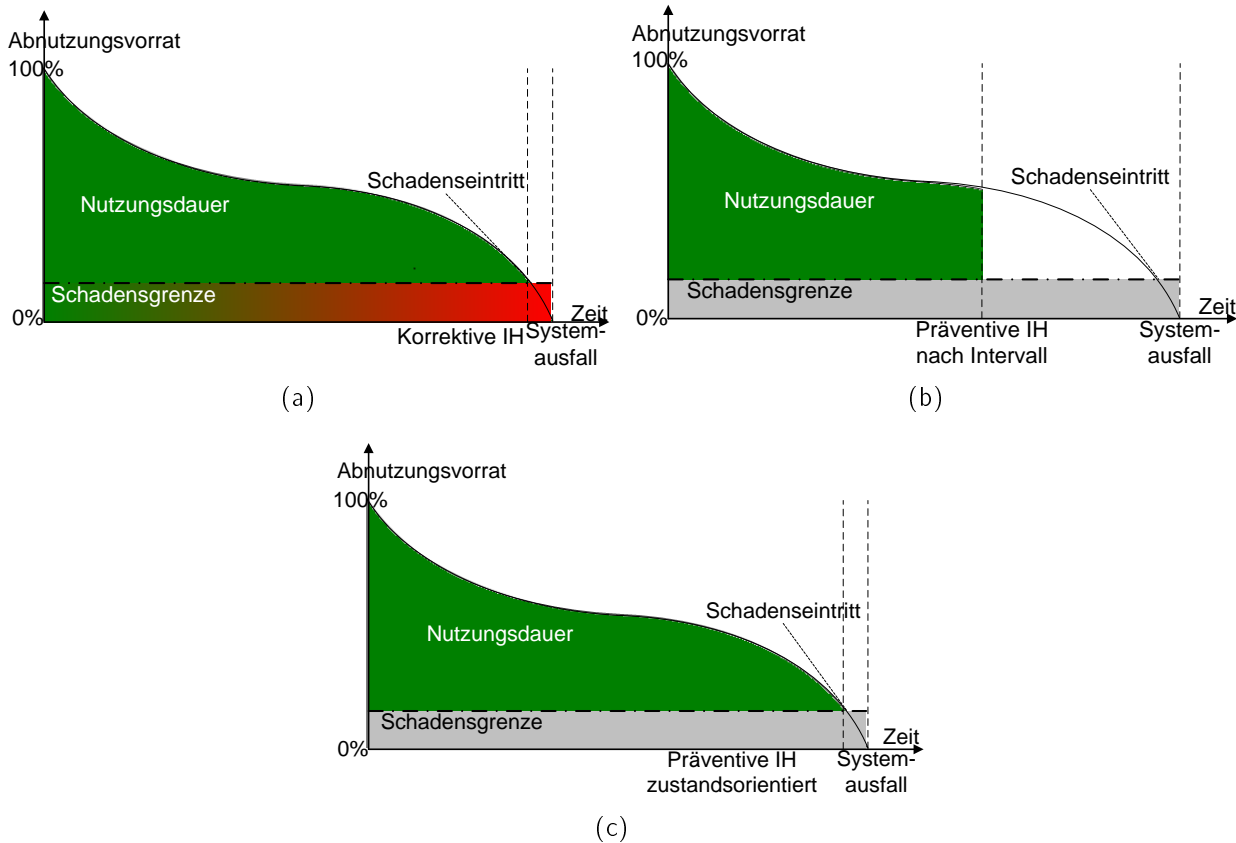


Abbildung 2-10: Verlauf des Abnutzungsvorrates und der Nutzungsdauer bei korrekativer Instandhaltung (a), präventiver Instandhaltung mit fixem Zeitintervall (b) und präventiver zustandsorientierter Instandhaltung (c) (in Anlehnung an [Böh12c, S. 52])

Im Gegensatz dazu ist die **präventive Instandhaltung** vorbeugend. Hier wird Instandhaltung nach festgelegten Kriterien durchgeführt, bevor das Schadensereignis eingetreten ist. Dafür muss der Abnutzungsvorrat bzw. sein Verlauf mindestens punktuell bekannt sein, damit rechtzeitig reagiert werden kann. Präventive Instandhaltung wird sehr häufig angewendet, wenn aus einer Störung oder einem Fehler ein Sicherheitsrisiko erwächst oder wenn die Instandhaltungskosten der korrekativen Strategie zu hoch sind. Es wird zwischen **zeitabhängiger** und **zustandsorientierter** (präventiver) Instandhaltung unterschieden. Bei ersterer finden Maßnahmen nach einem vorherbestimmten Zeitintervall oder festgelegten Nutzungseinheiten statt, ohne dass der Zustand genauer inspiziert wird. Dadurch sind Maßnahmen gut planbar. Allerdings kommt es häufig zu Effizienzverlusten durch ungenutzten Abnutzungsvorrat (siehe Abbildung 2-10(b)) und seltener auch zu Schadensereignissen durch beschleunigte Abnutzung zwischen den Instandhaltungsintervallen. Im Unterschied dazu wird mit einer zustandsorientierten Instandhaltung dem Wunsch nachgekommen, anhand von Zuständen der Anlagen notwendige Instandhaltungsmaßnahmen zeit-, qualitäts- und kostenoptimal, d.h. am echten Bedarf orientiert zu planen und durchzuführen (siehe Abbildung 2-10(c)). Voraussetzung dafür ist allerdings die kontinuierliche, zuverlässige Kenntnis über den Anlagenzustand.

In Kombination mit den Begriffen aus Abschnitt 2.2.1 lässt sich so auch die Zustandsprognose definieren.

Die **Zustandsprognose** ist eine Vorhersage der RUL oder einer äquivalenten Größe des Abnutzungsvorrats, zur rechtzeitigen Einleitung von Instandhaltungsmaßnahmen, mit dem Ziel, eine zustandsorientierte Instandhaltungsstrategie zu realisieren.

Die Wahl einer geeigneten Instandhaltungsstrategie wird allgemein als maßgebliches Mittel zur Optimierung der Instandhaltungskosten angesehen [Fis84, Ada89, BL04]. In den vergangenen zehn bis 20 Jahren hat sich in vielen Branchen, die Investitionsgüter mit ähnlichen Eigenschaften wie bei Weichen (siehe Abschnitt 2.2.2) betreiben, die zustandsorientierte Instandhaltung als optimale Strategie herausgestellt [Bie04]. Dabei beweist sie die Fähigkeit nicht nur die Instandhaltungskosten zu senken, sondern gleichzeitig auch die Verfügbarkeit zu erhöhen [Stu96, S. 107]. Nachteilig ist der hohe wissenschaftliche und technische Aufwand, der notwendig ist, um eine zuverlässige Zustandsbestimmung zu entwickeln und zu implementieren.

2.2.5 Schritte zur Umsetzung einer zustandsorientierten Instandhaltung

Aus den vorangegangenen Betrachtungen des Abschnitts 2.2 geht hervor, dass eine zustandsorientierte Instandhaltung für Weichen optimal wäre. Wie kann ein Anlagenbetreiber die zustandsorientierte Instandhaltung umsetzen, wenn er sich für diese Strategie entschieden hat? Das schließt auch die Frage ein, wie die Erkenntnisse dieser Arbeit in die Praxis überführt werden können. Eine Antwort darauf gibt der folgende Abschnitt anhand der Norm ISO 17359 [ISO03]. Sie definiert folgende Schritte zur Realisierung der zustandsorientierten Instandhaltung:

1. In einem **Anlagenaudit** legt der Betreiber zunächst fest, was die normale Funktion der Anlage ist und unter welchen Bedingungen sie betrieben werden soll.
2. Durch eine **Analyse der Kritikalität von Anlagen für den Betrieb und ihrer Zuverlässigkeit** priorisiert der Betreiber, welche Anlagen vorrangig zu behandeln sind. Dabei richtet sich die Priorisierung nach bestimmten Kriterien, z.B. den Kosten von Ausfällen, den Ausfallraten, den Auswirkungen auf den Gesamtbetrieb oder auf die Sicherheit.
3. Der Betreiber legt die **Messmethoden** fest, mit denen er die Anlage überwachen kann. Er bestimmt die Genauigkeit, den Ort und die Häufigkeit der Messungen. Außerdem wird der Zugang zu Messtechnik und die Durchführbarkeit der Messungen geprüft. Zusätzlich legt der Betreiber fest, welche Daten aufgezeichnet und verarbeitet werden.
4. Das **Sammeln und Analysieren von Messdaten** dient der eigentlichen Zustandsbeurteilung. Messwerte werden mit Sollwerten verglichen. Wiederkehrende Muster, Trends oder auffällige Abweichungen werden für die Diagnose und die Prognose herangezogen.
5. Auf Basis der Messdatenanalyse werden die passenden **Instandhaltungsmaßnahmen** eingeleitet.
6. Die Ergebnisse der bisherigen Schritte werden einem **Review** unterzogen und ggf. angepasst.

Der beschriebene Prozess sollte fortlaufend iterativ angewendet werden, da sich die Randbedingungen ständig ändern oder neue Technologie verfügbar wird. [ISO03]

Für Investitionsgüter mit einer langen Lebensdauer sind die Schritte eins und zwei meist schon während der Anschaffung oder Inbetriebnahme durchgeführt worden. Durch den Fortschritt, insbesondere bei der computergestützten Datenverarbeitung und -analyse, eröffnen sich auch für bereits seit 20 Jahren in Betrieb befindliche Anlagen neue Möglichkeiten. Moderne Sensoren und Diagnosesysteme sind in der Lage, Parameter, die über den Zustand Auskunft geben, präziser und häufiger zu erfassen als bisher. Es besteht Anpassungsbedarf bei den Schritten drei und vier. Böhm und Gutsche [BG11] schlagen daher ein etwas abweichendes Vorgehen vor, welches sich speziell auf schon länger betriebene Anlagen konzentriert und vom unerwünschten Zustand ausgeht – dem Ausfall.

Grundsätzlich werden Sensoren eingesetzt, um Parameter zu erfassen, aus denen sich direkt oder indirekt der Anlagenzustand bestimmen lässt. Zum Beispiel gibt eine Ultraschallmessung der Schiene direkt Auskunft über einen Materialfehler. Eine Messung der Wirkleistung des Weichenantriebsmotors lässt hingegen nur indirekt einen Rückschluss auf den Weichenzustand zu. Die Messdaten der Sensoren sind die Basis für ein Diagnose- bzw. Prognosesystem, welches die Daten interpretiert und so eine Beurteilung des Istzustands oder des zukünftigen Zustands durchführt. [BG11] Die Abbildung 2-11 veranschaulicht diesen Prozess.

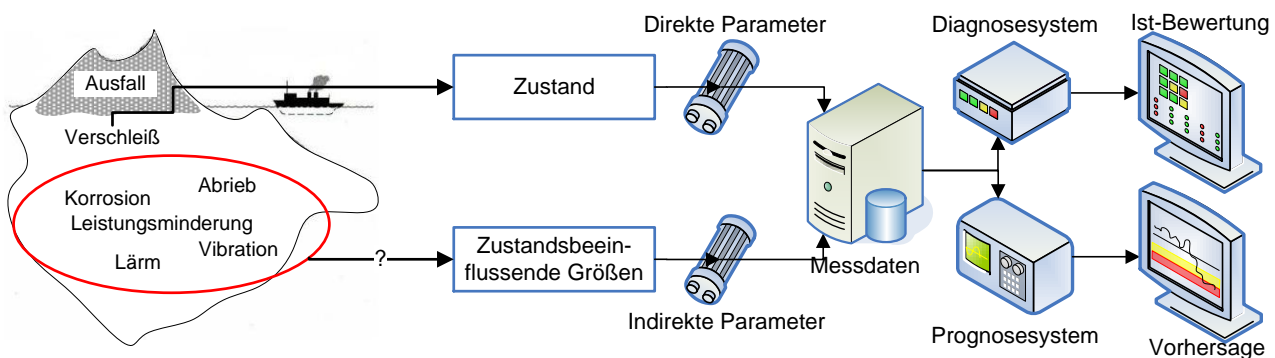


Abbildung 2-11: Prozess der Diagnose und der Prognose des Zustands (in Anlehnung an [BG11])

2.3 Zustandsorientierte Instandhaltung bei der Eisenbahninfrastruktur

Im Anschluss an den Abschnitt 2.2, in welchem allgemein dargelegt wurde, warum die zustandsorientierte Instandhaltung für Investitionsgüter als optimal in Bezug auf Kosten und Sicherheit gilt, wird nun die Situation bei der Eisenbahninfrastruktur betrachtet. Dass die zustandsorientierte Instandhaltung in vielen Bereichen der Eisenbahninfrastruktur als beste Strategie angesehen wird, ist schon länger bekannt. Den Nachweis erbrachte die Studie InfraCost, die im Auftrag des UIC erstellt und 2002 abgeschlossen wurde [BR02]. Die DB Netz AG setzte sich bereits 2002 das Ziel, die Verfügbarkeit zu erhöhen und die LCC zu senken [Kne02, S. 29]. Geeignetes Mittel dafür sollte der „...stufenweise Übergang von Fristinspektionen zur zustandsorientierten Instandhaltung“ [Kne02, S. 33] sein. Zwar wird die deutsche Infrastruktur präventiv instand gehalten, das muss sie wegen der Sicherheitsanforderungen auch [Fri09], aber eine zustandsorientierte Strategie mit kontinuierlicher Überwachung ist kaum Praxis. Daher blieben die Ziele Senkung der LCC und Verbesserung der Pünktlichkeit als Teil der integrierten Technologiestrategie auch 2009 hochgradig relevant [Bus09]. Der folgende Abschnitt legt dar, warum diese Ziele weiterhin priorisiert werden müssen und welcher Handlungsbedarf sich für ihr Erreichen ergibt.

2.3.1 Handlungsbedarf bei der Umsetzung einer zustandsorientierten Instandhaltung

In ihren Richtlinien 821 [DB 10], 824 [DB 09] und 892 [DB 02] legt die DB Netz AG ihre Instandhaltungsstrategie und -maßnahmen detailliert fest. Einem zustandsorientierten Ansatz soll darin mit belastungsabhängigen Fristen Rechnung getragen werden. Einerseits ist die Instandhaltung dadurch sehr gut planbar. Andererseits ist die Grundlage der Fristeinteilung zu grob gefasst. An lediglich drei Belastungskategorien (schwach, normal und stark) werden die Zeitintervalle für Inspektionen oder Wartungsaktivitäten ausgerichtet. Die Belastungskategorien hängen von der Streckenkategorie sowie der angenommenen Streckenbelastung ab und werden jährlich im voraus bestimmt. Dass damit eine gute oder gar optimale Orientierung der Instandhaltung am Anlagenzustand erfolgt, wurde von Beck et al. widerlegt [BJL07]. Bei einem einfachen Untersuchungsszenario mit einem einzigen Zugtyp (Gesamtgewicht 500 Tonnen), ergibt sich innerhalb der mittleren Belastungskategorie eine große Spannweite. So kann die Zahl jährlicher Zugfahrten zwischen 14235 und 57670 liegen. Die betreffende Infrastruktur wäre trotzdem nach denselben Intervallen zu inspizieren und zu warten. [BJL07, S. 9] Bei der Strategie der DB Netz AG handelt es sich also der Definition nach (siehe Abschnitt 2.2.4) eher um eine zeitabhängige Strategie. Bei genauerer Zustandsbeurteilung könnten laut Beck et al. bei den untersuchten Szenarien zwischen 79 und 87 Prozent der Instandhaltungskosten eingespart werden [BJL07, S. 10].

Weiterhin bietet gerade die zustandsorientierte Instandhaltung eine Lösung, den wachsenden Anforderungen an das Schienennetz zu begegnen. *„Wegen des zunehmenden Verkehrsaufkommens muss die Leistungsfähigkeit der Infrastruktur gesteigert werden“* [DB 12a, S. 27], selbstverständlich bei gleicher Sicherheit. Mehr Verkehrsbelastung bedeutet schnellere Abnutzung der Infrastruktur, wodurch sich folglich die Instandhaltungskosten erhöhen. Dies zeigen sowohl theoretische Betrachtungen (z.B. in Böhm et al. [BBKJ08] oder Quiroga et al. [QWS09]), als auch die Praxiserfahrungen mehrerer Europäischer Infrastrukturbetreiber (z.B. in Gesquière und Gardin [GG10]). Die Ergebnisse in Böhm et al. bestätigen, dass bei der LST eine bessere Orientierung am Anlagenzustand die Instandhaltungskosten reduzieren und die Verfügbarkeit erhöhen kann [BBKJ08].

Angesichts einiger Kennzahlen zur Instandhaltung der DB Netz AG (siehe Abbildung 2-12) wird der Handlungsbedarf deutlich. Im Vergleich zu 2009 haben sich die Verspätungen deutlich erhöht, ebenso sind die Störungsdauer (Störbestehenszeit) und die Instandhaltungsaufwände gestiegen. Lediglich die Anzahl an Störungen ist aufgrund von zurückgebaute Infrastruktur gesunken.

Von dem in Abschnitt 2.2.5 genannten Vorgehen zur Realisierung der zustandsorientierten Instandhaltung, müssen die Schritte eins und zwei nicht durchgeführt werden. Es ist bestens bekannt, unter welchen Bedingungen die Infrastruktur wie funktionieren soll. Auch die kritischen Elemente sind identifiziert. Nach Ansicht von Schilling und Lücking [SL03], der Studie Infra-Cost [BR02], Rama und Andrews [RA13] und vor allem von Rausch [Rau10] ist es die Weiche. Mit 19 Prozent Anteil an den infrastrukturbedingten Störungen bilden Weichen den Störungsschwerpunkt in Deutschland (siehe Abbildung 1-2 bzw. Abschnitt 1.2). Die Richtlinien 821, 824 und 892 [DB 10, DB 09, DB 02] definieren die Instandhaltungsmaßnahmen sehr detailliert. Die Maßnahmen müssen für eine Umsetzung der zustandsorientierten Strategie nicht überarbeitet werden. Vielmehr fehlt die Voraussetzung, d.h. die genaue Kenntnis des Verlaufs des Abnut-

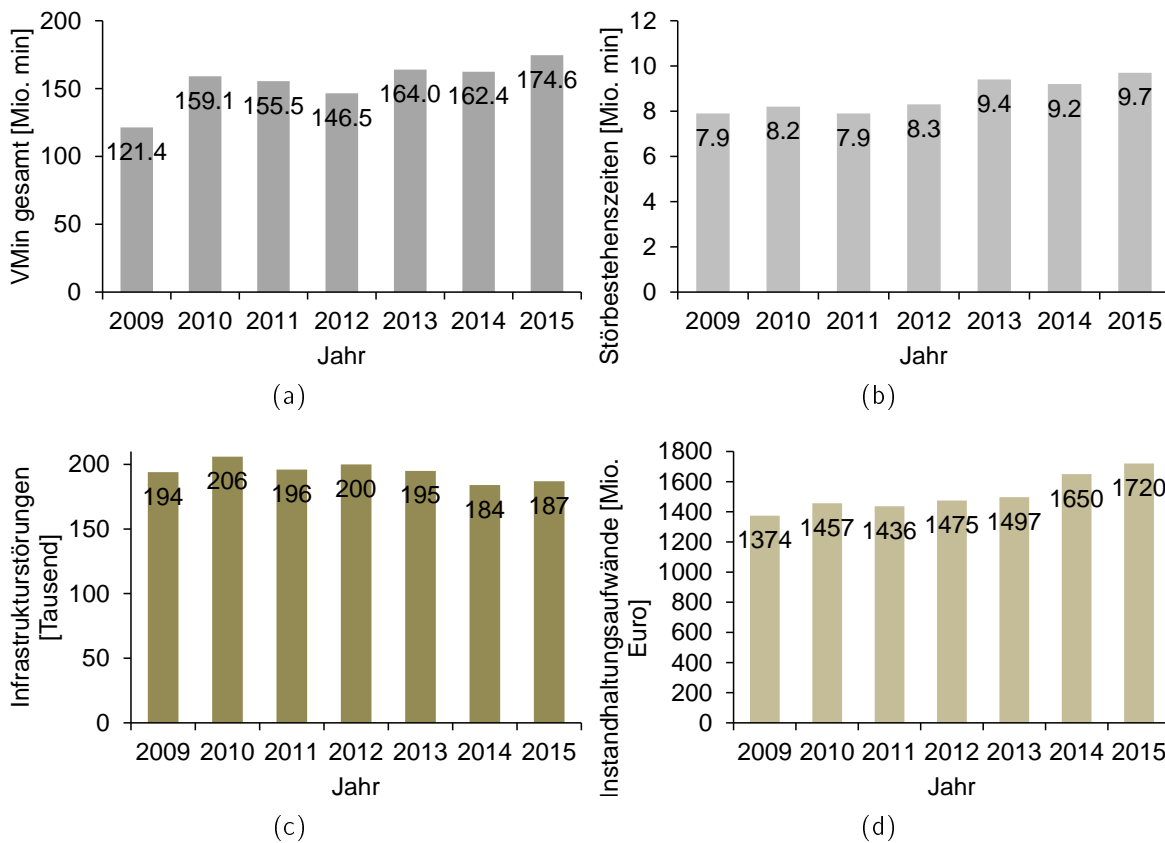


Abbildung 2-12: Entwicklung der Verspätungsminuten (a), der Störbestehenszeiten (b), der Gesamtstörungszahl (c) und der Instandhaltungsaufwände (d) (Zahlen nach [DB 14b, DB 16])

zungsvorrates oder eine kontinuierliche Zustandsüberwachung. Eine manuelle Inspektion für die Zustandsbeurteilung ist bei den aktuell 67071 Weichen und Kreuzungen im Netz der DB AG [DB 16] undenkbar. Die derzeit wöchentlichen Inspektionen (bei höchster Belastungskategorie) wären mehrmals täglich durchzuführen. Gemäß des Vorschlags von Böhm und Gutsche [BG11] sollten deshalb Sensoren eingesetzt werden, die die relevanten Daten erfassen. Sensoren, die im Gleisbett eingesetzt werden sollen, müssten sehr robust gegenüber Schmutz, Erschütterungen, etc. sein. Das macht sie meist auch teuer. Gleichzeitig ist jeder Sensor per se eine Fehlerquelle, die ausfallen kann und Wartung benötigt. Darum ist aus wirtschaftlichen und technischen Gründen darauf zu achten, so wenig Sensoren wie möglich einzusetzen. Die Auswertung der Sensordaten sollte Diagnose- bzw. Prognosesystemen überlassen werden, da die zu erwartende Datenmenge kaum mehr durch den Menschen handhabbar ist. [BG11] Statistik und Data Mining halten hierfür geeignete Werkzeuge bereit. [Böh13a, Böh13d, GKVJ13].

2.3.2 Systeme zur Fernerfassung und Überwachung des Weichenzustands

Im vorherigen Abschnitt wurde anhand der Kennzahlen zu Verspätungsminuten (VMin), Störbestehenszeiten, Störungsanzahl und Instandhaltungsaufwänden gezeigt, dass die von der DB Netz AG angestrebte zustandsorientierte Instandhaltung für Weichen noch nicht ausreichend realisiert wurde. Es bedarf hauptsächlich einer kontinuierlichen Zustandsüberwachung mittels Sensorik und einer Zustandsvorhersage. Der folgende Abschnitt geht der Frage nach, ob bereits Überwachungssysteme am Markt verfügbar sind, die diese Aufgabe übernehmen könnten.

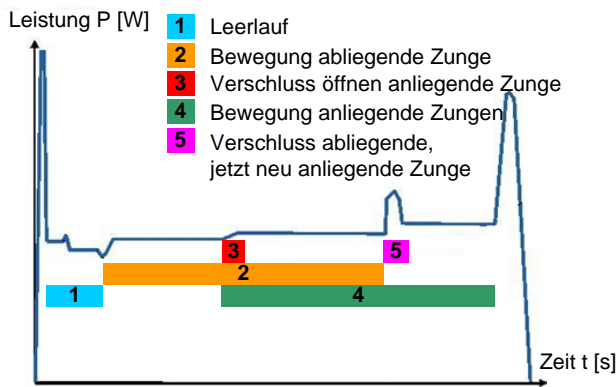
Die DB Netz AG setzt an vielen verschiedenen Stellen im Netz Diagnosesysteme ein. Die Überwachungssysteme erfüllen jedoch nicht die Ziele des Betreibers.

„Zumeist stellen die Diagnosen im Störfall dem Instandhaltungspersonal Zustandsdaten der Anlage zur Verfügung, um die Entstörung zu beschleunigen. Der gestörte Zustand der Anlage soll aber gar nicht eintreten und das eigentliche Ziel der Zustandsüberwachung und einer frühzeitigen Meldung wird nur selten erfüllt.“ [Hol12, S. 20]

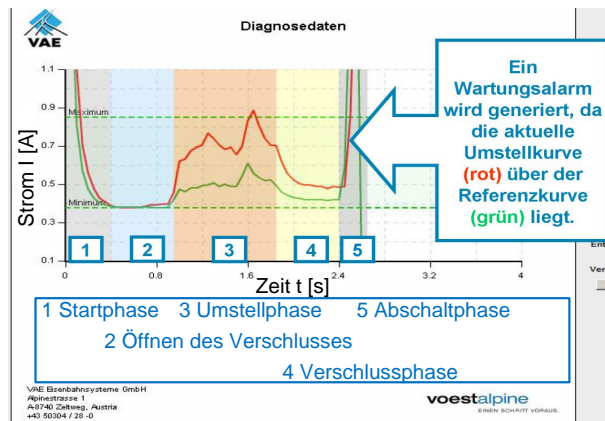
Auch für Weichen existieren Diagnosesysteme, genauer gesagt für den Antrieb. In Deutschland sind derzeit zwei Produkte für die Überwachung von Weichenantrieben bei Eisenbahnen im Sinne des AEG zugelassen. Das ist zum einen das System SIDIS W (compact) der Siemens AG und zum anderen das System VAE Roadmaster 2000 der voestalpin AG. Letzteres ist in den drei Ausprägungen Light, Advanced und Pro erhältlich. Während bei der Variante Light eine Antriebsüberwachung nur aus dem Stellwerk heraus erfolgt, sind in der Advanced Version zusätzliche Sensoren an der Weiche verbaut. Bei der Variante Pro werden darüber hinaus signaltechnische Anlagen überwacht. In der Weichenüberwachung unterscheiden sich Advanced und Pro jedoch nicht. [Sei08] Nach Aussage der DB Netz AG setzt sie nur den Roadmaster 2000 Light ein. Bei einer Pilotanwendung hatte sich die Advanced-Version nicht bewährt, weil die Sensoren weitere Fehlerquellen darstellten und aus den zusätzlich erfassten Parametern kein direkter Nutzen ersichtlich war [Gut08]. Daher wird an dieser Stelle lediglich auf SIDIS W und Roadmaster 2000 Light eingegangen.

Beide System funktionieren im Grunde sehr ähnlich. Sie beruhen auf dem Ansatz, dass der Stromverbrauch des Antriebs proportional zur Leistung ist, die der Motor für den Umstellvorgang erbringen muss [BS06]. Fehler, wie mangelnde Schmierung der Gleitstühle oder der Verschlüsse, verklemmte Verschlüsse, verbogenen Zungen oder Blockierung durch Fremdkörper erhöhen die benötigte Leistung und sind daher theoretisch durch eine Stellstromüberwachung detektierbar. Beide Systeme erfüllen eine wichtige, sicherheitsrechtliche Voraussetzung für den operativen Einsatz. Sie sind völlig rückwirkungsfrei in Bezug auf die Sicherungstechnik bzw. die Fahrweigestellung, weil sie die benötigten Information über einen induktiven Sensor erfassen. Dazu wird eine Klemme (Durchsteckstromwandler) um das Kabel gelegt, welches den Stellbefehl vom Stellwerk zur Weiche leitet. Während SIDIS W die elektrische Stromstärke (I) und die elektrische Spannung (U) misst und daraus die Wirkleistung (P) des Motors berechnet, verwendet Roadmaster nur die Stromstärke. Wie in Abbildung 2-13 zu sehen ist, zeigt das idealisierte Messsignal bzw. die Referenzkurve bei beiden Fällen einen ähnlichen Verlauf in Abhängigkeit von der jeweiligen Umstellphase (Erläuterung zum Umstellvorgang siehe Abschnitt 2.1.3).

Der wesentliche Unterschied zwischen SIDIS W und Roadmaster besteht in der Interpretation der Messungen für die Fehlerdetektion. SIDIS W berechnet für ausgewählte Bereiche jeder Umlaufphase jeweils den Durchschnittswert der Wirkleistung. Roadmaster hingegen bildet lückenlos das Integral der Stromstärke in jeder Umlaufphase. Aber beide Systeme vergleichen diesen Wert mit einem Referenzwert, der bei der Inbetriebnahme oder der Wartung aufgezeichnet wird. Weicht der Durchschnitt respektive das Integral um einen konfigurierbaren Prozentsatz von der Referenz ab, wird ein Alarm angezeigt. Alarme werden in zwei Warnstufen unterschiedlicher Dringlichkeit signalisiert, gelb und rot. Der Hersteller Siemens empfiehlt bei Gelb eine baldige Wartung,



(a) Idealisierter Verlauf in Anlehnung an [SB02]



(b) Beispielhafter Verlauf in Anlehnung an [Win10]

Abbildung 2-13: Messkurvenverlauf der Diagnosesysteme SIDIS W (a) und Roadmaster 2000 Light (b) während der Umstellphasen einer Weiche

während bei Rot eine sofortige Wartung notwendig ist (vgl. [KR11, S. 7]). Ganz ähnlich deutet voestalpine die Alarmer, Gelb als Zustandsalarm und Rot als Anlagenfehler (vgl. [Win10, S. 16]). Das Produkt von Siemens bringt eine proprietäre Software inklusive Bedienoberfläche mit. Im Gegensatz dazu kommuniziert voestalpine die Messdaten über Standardwebschnittstellen, womit die Bedienung über einen gewöhnlichen Internet-Browser einhergeht. In dieses Konzept passt, dass Daten beim Roadmaster im XML-Format, aber auch als MS Excel Dateien exportiert werden können. SIDIS W bietet zu Datenexportzwecken MS Access Dateien an. In beiden Systemen kann der Nutzer sich die errechneten Durchschnittswerte bzw. Integrale mehrerer Umläufe ansehen. Er muss dabei aber selbst einen Trend interpretieren. Tabelle 2-1 fasst die Eigenschaften beider Systeme zusammen.

Das System SIDIS W ist zwar seit 2011 in einer überarbeiteten Version unter der Bezeichnung SIDIS W compact erhältlich. Seine Diagnosefähigkeiten und Eigenschaften haben sich allerdings kaum verändert, abgesehen von einer vereinfachten Anzeige und der Reduktion der Durchschnittswerte auf weniger Umlaufphasen (vgl. [KR11]). In den Artikeln zum SIDIS W (compact) von Stoll und Bollrath [SB02], Bollrath und Sattler [BS06] sowie Körkemeier und Robbe [KR11] wird zwar beschrieben, welche Fehler und Störungen mit Hilfe des Systems potenziell diagnostiziert werden könnten, aber eine explizite Diagnose gibt das System nicht ab (vgl. [Bol07]). Ähnliches gilt für Roadmaster und die Veröffentlichung von Winter [Win10]. Mit ihren Alarmen sind SIDIS W und Roadmaster weniger Diagnose- als vielmehr Detektionssysteme. In dieser Funktion zeigen sie noch einiges Verbesserungspotenzial, wie zumindest für SIDIS W durch die statistischen Analysen von Böhm [Böh12a, Böh12c] nachgewiesen wurde. Bei knapp 295000 Umläufen wurden rund 37000 rote Alarmer (etwa 12.5 Prozent) signalisiert. Da tatsächlich nur 84 Störungen auftraten, gab es deutlich zu viele Fehlalarme [Böh12c].

Die beschriebenen Diagnosesysteme bei der DB AG oder ähnliche Systeme im Ausland befinden sich zwar im Feld, doch verlässliche Prognosen bieten sie nicht. Ihre Überwachungsdaten können jedoch ein wesentlicher Baustein für die Vorhersage sein, wie im Verlauf der Arbeit noch ausführlich erläutert wird (siehe Abschnitt 5.1). Dass diese Weichendiagnosesysteme zur Reduktion der LCC beitragen, haben García Márquez et al. am Beispiel Englands nachgewiesen [GMLTR08]. Der Erfolg beruht allerdings nicht auf den Systemen selbst, sondern auf den

Diagnosesystem	Siemens AG	Voest Alpin AG
Eigenschaft	SIDIS W	Roadmaster 2000 Light
Funktionsweise	rückwirkungsfreie Messung elektrischer Signale mittels induktivem Sensor am Übertragungskabel der Stellinformation zum Antrieb	
Primäre Messgröße des Zustands	durchschnittliche Wirkleistung P in Ausschnitten verschiedener Umlaufphasen	Integral der Stromstärke I während verschiedener Umlaufphasen
Zustandsinterpretation	Vergleich der primären Messgröße mit einer Referenz und Anzeige des Status (grün, gelb oder rot) je nach Höhe der prozentualen Abweichung von der Referenz	
Trendanalyse	Anzeige einzelner primärer Zustandsgrößen für mehrere Umläufe, augenscheinliche Trendinterpretation durch Nutzer	
Störungsdetektion	durch Zustandsinterpretation und Alarmer	
Störungsdiagnose	erfolgt nicht, da keine Störungsursache ermittelt wird	
Störungsprognose	erfolgt nicht, da keine RUL oder Abnutzungsvorrat ausgegeben wird	
Anzahl insgesamt gemessener Parameter	56 (Auflistung siehe Tabelle B-1 in Anhang B.2)	mindestens 11 (genauere Angaben liegen nicht vor)
Messfrequenz [Hz]	50	20
Datenspeicherung	Server	Datenbankserver
Datenzugriff	proprietäre Schnittstelle	Standardwebschnittstelle
Visualisierung und Bedienoberfläche	proprietäre Software	webbasiert über Internet-Browser
Datenexport	MS Access	XML, MS Excel
Verbreitung bei der DB AG (Stand: Juni 2011)	117 Antriebe im Regionalbereich Ost, unbekannte Anzahl bei S-Bahn München	circa 20 Antriebe im Regionalbereich Ost
Verbreitung bei der DB AG (Stand: Jan 2013)	über 130 Antriebe in mehreren Regionalbereichen, unbekannte Anzahl bei S-Bahnen München und Frankfurt	circa 86 Antriebe in mehreren Regionalbereichen

Tabelle 2-1: Eigenschaften der Weichendiagnosesysteme SIDIS W und Roadmaster 2000 Light

Forschungsergebnissen zur automatischen Fehlerursacheninterpretation der Überwachungsdaten (vgl. [ZAB⁺02, GMWR07, CSAA08, Sil09, ACBS09, Bai10, ECK10, ARK13]). Dadurch lässt sich zwar die Entstörung effizienter gestalten, d.h. die Störungsdauer verringern. Verhindern lassen sich Störungen damit aber kaum. So erklärt sich, dass die integrierte Investitions- und Instandhaltungs-Strategie („3-i“) der DB Netz AG, gestartet 2006, auch vier Jahre später noch zu dem Schluss kommt: *„Die Erhöhung der Planbarkeit von (Instandhaltungs-) Maßnahmen ist der Schlüssel dafür, Qualität und Verfügbarkeit der Anlage weiter zu steigern, die Lebensdauer zu erhöhen und eine material- und ressourcenoptimierte Instandhaltung zu gewährleisten. Es gilt der Grundsatz: Agieren statt reagieren.“* [Gra10, S. 169]. Das gelingt allerdings erst durch eine verlässliche Zustandsprognose.

2.4 Bestehende Prädiktionsmodelle und ihre Anwendbarkeit für Weichen

Bisher wurden in diesem Kapitel die wesentlichen Grundlagen erläutert. Dazu wurden der Aufbau und die Funktion einer Weiche detailliert beschrieben (Abschnitt 2.1) sowie der Zusammenhang

zwischen LCC und Instandhaltungsstrategie erklärt (Abschnitt 2.2). Der Abschnitt 2.3 legte daraufhin dar, dass in der Praxis insbesondere eine Störungsvorhersage bei Eisenbahnweichen fehlt, um die zustandsorientierte Instandhaltung umzusetzen und ihre Vorteile zu nutzen. Da theoretisch zahlreiche Methoden für die Prädiktion existieren, stellt sich die Frage, inwieweit diese schon für eine Zustandsprognose bei Eisenbahnweichen angewendet wurden. Der folgende Abschnitt beantwortet diese Frage und diskutiert den Stand der Forschung. Bestehende Ansätze werden danach beurteilt, ob sie für eine präzise, verlässliche und kostensenkende Prognose von Weichenstörungen im Netz der DB AG geeignet sind. Genauer bedeutet dies:

- Wurde die Prognose an mehreren Weichenobjekten durchgeführt und besitzt so eine gewisse Allgemeingültigkeit?
- Weisen die Untersuchungsobjekte analoge Eigenschaften zu Objekten bei der DB AG auf, so dass die Erkenntnisse übertragbar sind?
- Basiert eine Prognose auf realen Daten?
- Wird eine RUL oder ähnliche Größe prognostiziert?
- Lässt diese Prognose ausreichend Eingreifzeit, um die Störung zu verhindern?

Wie in der vorliegenden Arbeit bereits argumentiert wurde (siehe Abschnitte 1.2 und 2.3), wird die Zustandsdiagnose dabei nicht betrachtet, da sie keine vorausschauende Störungsprävention ermöglicht. Einen guten Überblick über die bisherige Forschung zur Zustandsdiagnose bei Eisenbahnweichen geben García Márquez et al. [GMRT10] und Vileiniskis et al. [VRPRA13] sowie etwas ausführlicher die Dissertationsschriften von Silmon [Sil09] und Asada [Asa13].

Eine der frühesten Arbeiten zu Zustandsbeurteilung stammt aus dem Jahr 2004 von Pedregal et al. [PGMS04]. Mit der Absicht bestimmte Fehlertypen des Weichenantriebs zu diagnostizieren, wurde ein experimentelles Szenario aufgebaut. Darin wurde ein einzelner Weichenantrieb unbekannten Typs mit Sensoren ausgerüstet. Diese erfassen die elektrische Stromstärke (I) am Motor und die Stellkraft (F) an der Schieberstange. Allerdings verwendeten die Forscher hauptsächlich das Kraftsignal. Anschließend simulierten sie verschiedene, nicht näher bezeichnete Fehlertypen. Die Wissenschaftler entwickelten anhand von 151 Umlaufmessungen ein bivariates Komponentenmodell des Zustandsraumes, bestehend aus Signal und Rauschen. Dieses Modell nutzten sie zur Vorhersage des Stellkraftsignals innerhalb des aktuellen Umlaufs, um zu entscheiden, ob es sich um einen fehlerfreien oder einen fehlerhaften Umlauf handelt. Eine Trendvorhersage der Stellkraft als Zustandsindikator erfolgt jedoch nicht über mehrere Umläufe. In der 2009 erschienen Veröffentlichung zur wiederholten Anwendung derselben Methode wichen die Resultate kaum von den vorangegangenen Ergebnissen ab [PGMR09]. Diesmal nutzte das Forscherteam reale Messdaten von drei elektromechanischen Antrieben des Typs M63 und vier elektrohydraulischen Antrieben aus dem Zeitraum von Mai bis November 2001. Auch hier erfolgt die Vorhersage des aktuellen Umlaufsignals zu Beginn des Umlaufs. Eine Anwendung für mehrere Umläufe wird nicht erwähnt, geschweige denn eine Angabe des Abnutzungsvorrates oder der RUL. Das Modell ist somit nicht als Zustandsprognose im Sinne der Definition in Abschnitt 2.2.4 nutzbar.

In seiner Dissertation zur Fehlerdiagnose aus dem Jahr 2009 verwendete Silmon den Weichenantrieb als ein Fallbeispiel [Sil09]. Er analysierte experimentell simulierte Fehlerzustände von drei elektromechanischen Antrieben des Typs HW. Seine Datenbasis bildeten die Stromstärke am

Motor und die Kraft an der Schieberstange während der Weichenumläufe. Sein Ansatz bestand darin, die Messkurve in bestimmte Abschnitte (Episoden) zu zerlegen. Jede Episode wurde, gemäß ihres qualitativen Kurvenverlaufs, mit einem Buchstaben von A bis I codiert, z.B. A für einen annähernd horizontalen Verlauf, B für einen ungefähr exponentiellen Anstieg, C für einen eher linearen Anstieg, etc. Damit konnte er jeden Umlauf durch ein Wort abbilden. Die Zuordnung der Worte zu einem von drei Zuständen (funktionierend, zu hohe Stellkraft oder zu niedrige Stellkraft) realisierte Silmon mit Fuzzy Regeln. Neben der gelungenen Diagnose der simulierten Fehler, schlug er auch lineare Regression zur Trendberechnung jedes Zustands und der linearen Kombination aller drei Trends vor. Dadurch sollte die Stärke des Vorhandenseins eines Fehlers im Intervall von 0 bis 1 ausgedrückt werden können und somit auch eine Interpretation als Zustandsprognose. Dies gelang jedoch nicht, abgesehen von dem Umstand, dass die lineare Regression für maximal 15 Umläufe eine unzuverlässig kleine Menge darstellte. So kam Silmon selbst zu dem Schluss, dass insbesondere die qualitative Trendanalyse deutlich verbessert werden muss [Sil09, S. 124 und S. 126]. Für eine Zustandsvorhersage im Sinne dieser Arbeit ist Silmons Ansatz daher ungeeignet.

Ein stochastisches Modell setzten Chattopadhyay und Kumar [CK09] in ihrer 2009 veröffentlichten Arbeit zur Prognose der Schienenalterung ein. Sie analysierten den statistischen Zusammenhang von je sechs Schienenbrüchen zweier Schientypen zur Belastung, gemessen als kumulierte Millionen Britische Tonnen (engl. Million Gross Tonnes) (MGT) beim Auftreten des Bruchs. Auf Basis realer Daten der schwedischen Bahn Banverket ermittelten sie die Parameter einer Weibull-Verteilung. Damit waren sie in der Lage, die mittlere Zeit bis zu einem Ausfall (engl. Mean Time To Failure) (MTTF) als Basis für eine Alterungsprognose auszuweisen. Die Übertragung der Ergebnisse ist allerdings nicht möglich, da die untersuchten Schientypen im Netz der DB AG nicht vorhanden sind und auch die Menge an Brüchen noch keine allgemeingültigen Aussagen zulassen, wie die Autoren selbst schrieben [CK09, S. 128]. Der Nutzen wäre zudem eher gering, weil Schienenbrüche sehr selten Ursache von Weichenstörungen sind.

An dieser Stelle sei erwähnt, dass Schienenfragestellungen grundsätzlich nur begrenzt für die Prognose geeignet sind und damit nur einen Bestandteil einer Weichenstörungsprognose bilden können. Bei den meisten westeuropäischen Bahnen herrschen ähnliche Verhältnisse. So zeigen z.B. die Zahlen der EcoSwitch Studie aus dem Jahr 2002, dass durchschnittlich nur 30 Prozent der Weichenstörungen in den Niederlanden auf die Schienen und Zungen zurückzuführen sind, während 50 Prozent dem Antrieb zugeschrieben werden [JZ02, S. 4]. Ähnliches haben auch Rama und Andrews in ihrer Veröffentlichung von 2013 bestätigt [RA13]. Die Autoren haben 2932 Weichenstörungen ausgewertet, die in Großbritannien über einen Zeitraum von 10 Jahren erhoben wurden. In 58 Prozent der Fälle war der Antrieb samt Verschluss die Störungsursache.

Zwanenburg beschäftigte sich in seiner 2009 veröffentlichten Dissertation explizit mit der Modellierung des Abnutzungsprozesses bei Weichen und Kreuzungen [Zwa09]. Er verwendete gleichfalls die kumulierte Belastung und zusätzlich die Liegedauer von mehr als 1736 Weichen, die er der Datenbank der Schweizer Bahn SBB entnahm. Für die Erstellung seines Zuverlässigkeitsmodells verwendete er eine robuste, lineare Regression multipler Parameter. Dadurch berücksichtigte er Eigenschaften wie den Weichenradius, die Unterbauqualität, die Befahrerichtung, die Befahrgeschwindigkeit und den Anteil an Güterzugüberfahrten, weil er diese als relevant für den Alterungsprozess ermittelt hatte. Seine Arbeit konzentrierte Zwanenburg allerdings auf die Prognose des

Ersatzzeitpunktes einer Weiche. Obendrein schloss er den Weichenantrieb komplett von seinen Untersuchungen aus [Zwa09, S. 12]. Das lineare Regressionsmodell erreichte in seiner Evaluation keine einzige Übereinstimmung zwischen prognostiziertem Ersatzzeitpunkt und tatsächlichem Ersatzzeitpunkt. Damit ist das Modell ungeeignet für die Prognose einer Weichenstörung.

In dem 2010 erschienenen Artikel von Holzfeind und Hummitzsch [HH10] wurde das Alterungsverhalten von Schienen mit polynomieller Regression beschrieben, um die Restlebensdauer zu prognostizieren. Dabei bezogen die beiden Forscher verschiedene Parameter mit ein, bspw. die Schwellenart, den Schotter, den Unterbau, die Belastung und die Instandhaltungsmaßnahmen. Sie nutzten dafür umfangreiche Aufzeichnungen der österreichischen Bahn ÖBB von über 4000 km Strecke aus dem Zeitraum 2001 bis 2009. Obwohl die Ergebnisse sehr gut waren, kann die Arbeit keinen Beitrag zur Weichenzustandsprognose liefern. Die Autoren schrieben selbst: *„Störstellen, wie Weichen oder Brücken, werden vorerst aufgrund eines erwarteten stark abweichenden Verhaltens aus den Betrachtungen ausgeschlossen.“* [HH10, S. 33]

Die Veröffentlichung von García Márquez et al. [GMPR10], die 2010 erschienen ist, beschäftigte sich mit dem Weichenantrieb. Darin verwendete das Team das autoregressiver, gleitender Durchschnitt (engl. Autoregressive-Moving-Average) (ARMA)-Verfahren mit bivariatem Eingang (Signal und Rauschen) zur Prognose der Umlaufzeit für den nächsten Weichenumlauf. Eine gewisse Über- oder Unterschreitung der Umlaufzeit galt den Autoren als Störungsindikator. Von insgesamt 760 Umläufen drei elektromechanischer M63 und vier elektrohydraulischer Antriebe lagen Daten zur Stromstärke des Motors oder der Pumpe, zur Kraft am Stellschieber, zum Öldruck und zum Ölstand vor. Diese Aufzeichnungen stammten aus einem britischen Bahnhof aus dem Zeitraum Mai bis November 2001. Die jeweils 50 letzten Messungen vor dem Zeitpunkt t wurden verwendet, um die Umlaufzeit für den Zeitpunkt $t + 1$ vorherzusagen. Wich diese prognostizierte Umlaufzeit zu sehr von der erwarteten Umlaufzeit im fehlerfreien Zustand ab, wurde ein Alarm signalisiert. Dadurch konnten die Autoren zwar eine gute Zustandsdiagnose erreichen, aber die Erwartung einer Prognose blieb unerfüllt. Da keine Analyse mit Vorhersage mehrerer zukünftiger Umläufe ($t + n$ mit $n > 1$) stattfand, lässt sich die Anwendbarkeit für RUL-Aussagen nicht prüfen.

Die Dissertation von Gutsche [Gut10], ebenfalls von 2010, ist zwar mit dem Ziel einer integrierten Bewertung von Investitions- und Instandhaltungsstrategien erstellt worden, enthält aber einen Ansatz für die Weichenzustandsbeurteilung. Die Autorin nutzte Betriebsdaten der DB Netz AG über die Anzahl der Umstellvorgänge und die Wirkleistung (P) von drei elektromechanischen Weichenantrieben des Typs S700K. Mit polynomieller Regression leitete sie aus circa 900 Umläufen ein Polynom dritten Grades zur Beschreibung der Abnutzung eines einzelnen Antriebs ab. Sie prognostizierte damit den Ausfallzeitpunkt in Abhängigkeit von den Umstellvorgängen. Da ihre Absicht nicht in der möglichst genauen Vorhersage des Störungszeitpunktes lag, sondern in der Entwicklung eines Verfahrens, mit dem sowohl die Investitions- als auch die Instandhaltungsstrategie bewertet werden können, nutzte sie das Polynom nur als Beispiel. So wurde schon aus der Betrachtung der verbliebenen zwei Antriebe deutlich, dass dafür völlig andere Koeffizienten gefunden werden müssten, wenn ein Polynom dritten Grades überhaupt passend wäre. So schrieb Gutsche selbst, dass ihre Gleichung nur ein Beispiel sei und der tatsächliche Ausfallzeitpunkt davon abweichen könne [Gut10, S. 113 f.]. Anders als im realen Betrieb, spielten Abweichungen von Tagen oder Wochen für ihren Betrachtungszeitraum von über 20 Jahren keine Rolle. Das

Beispiel eines einzelnen Antriebs ist kaum repräsentativ. Darüber hinaus sprechen die eigenen Erfahrungen gegen Regressionsmodelle zur Weichenzustandsprädiktion (vgl. [Böh12c]). Daher ist Gutsche's Ansatz nicht für die Prädiktion als Basis einer zustandsorientierten Instandhaltung nutzbar, obwohl sie eine Restlebensdauer in Form von Umstellvorgängen ausgibt.

Wegen des engen Bezugs zur Dissertation von Gutsche sei Folgendes ergänzend erwähnt: Im Rahmen der Forschung zur vorliegenden Arbeit wurde versucht, die Zeit bis zur nächsten Störung mit Hilfe einer linearen Regressionsfunktion multipler Parameter vorherzusagen. Dass dies nicht erfolgreich war, ist in Böhm [Böh12c] nachzulesen.

Von einem überwiegend türkischen Forscherteam stammt eine Reihe von Veröffentlichungen zur Zustandsprädiktion bei Weichen aus den Jahren 2010 und 2011. Sie alle sind, mehr oder weniger, vom selben Team, welches dieselbe Datenbasis verwendete. Dazu wurde ein elektromechanischer Antrieb unbekannten Typs mit Messsensoren für die lineare Position und die Kraft an der Schieberstange, für den Abstand zwischen Backenschiene und Zunge sowie für die Stromstärke und die Spannung am Motor ausgerüstet. Anschließend wurden bestimmte, simulierte Fehlerzustände gemessen und aufgezeichnet. Weiterhin wurde bei allen Analysen die Annahme zugrunde gelegt, dass der Verlauf der Abnutzung, also vom fehlerfreien zum gestörten Zustand, exponentiell wächst. Unter dieser Annahme unterteilten die jeweiligen Autoren den Verlauf in zehn diskrete Stufen und verwendeten ihn für ihre Prognosen bei folgendem Vorgehen:

- Yilboga et al. [YEGC10] verwendeten ein rekurrentes Neuronales Netz, um die Zuordnung der Messdaten und der vergangenen Zustände (eine der zehn diskreten Stufen) zum folgenden Zustand zu erlernen. Anschließend testeten sie das Netz an zehn Weichen, indem sie die RUL in Form von Umläufen vorhersagten und mit den tatsächlichen Verläufen verglichen. Erstaunlicherweise umfasste ihre Evaluation maximal eine Vorschau von 20 Umläufen. Das bedeutet, dass jede Weiche innerhalb von höchstens 20 Umläufen gestört war oder die Vorhersage nur für 20 Umläufe getestet wurde. Des Weiteren fällt auf, dass gerade am Anfang die Abweichung zwischen prognostizierten und tatsächlich verbleibenden Umläufen relativ hoch war. Sie lag bei ein bis sieben Umläufen von maximal 20, d.h. im schlechtesten Fall wurde nur eine Genauigkeit von 65 Prozent erreicht. In Anbetracht der Tatsache, dass viele Weichen im realen Betrieb 20 Umläufe in weniger als einer Stunde vollführen müssen, ist ein Vorhersagehorizont von 20 Umläufen nicht ausreichend, um eine Störung verhindern zu können.
- In einer zweiten Veröffentlichung aus dem Jahr 2010 nutzten Guclu et al. [GYE⁺10] ARMA mit denselben Daten, ebenfalls um die RUL, gemessen in Umläufen, vorherzusagen. In dieser Veröffentlichung beschrieben sie allerdings die Art der Störung und wie sie diese simuliert hatten. Als Störung wählten die Forscher eine Blockierung bzw. einen Schwergang durch mangelnde Schmierung der Gleitstuhlplatten, ein relevantes Problem im Regelbetrieb. Bei mangelnder Schmierung sorgt die Reibung der Zunge auf den Gleitstuhlplatten für einen erhöhten Kraftbedarf während des Umstellvorgangs. Ist der Reibungswiderstand zu hoch, kann die Weiche nicht mehr umgestellt werden. Um die Referenzdaten zu erhalten, wählten die Wissenschaftler ein rückwärtiges Vorgehen, vom gestörten zum fehlerfreien Zustand. Sie erfassten zunächst die Messwerte der völlig ungeschmierten Weiche. An dieser wurden dann nach und nach einzelne Gleitstühle geschmiert und jeweils die Messwerte erfasst, bis schließlich alle Gleitstühle ordentlich geschmiert waren. Die Parameter des

ARMA wurden mit diesen Daten bestimmt und anschließend an zehn Weichen evaluiert, analog zum vorhergehenden Artikel. Für maximal 20 Umläufe wurden die prognostizierten und die tatsächlichen Umläufe bis zur Störung verglichen. Die Ergebnisse fielen besser aus als bei Yilboga et al. [YEGC10]. Die Übereinstimmungen waren höher und die Abweichung zu Beginn betrug ein bis fünf Umläufe von maximal 20. Aber auch hier bleibt das Problem des eher kleinen Vorhersagehorizonts von weniger als 20 Umläufen bestehen.

- Eker et al. [ECG⁺11] veröffentlichten 2011 einen weiteren Artikel. Und wie zuvor schon Guclu et al. [GYE⁺10], analysierten sie dieselben Daten desselben Experiments. Ihre Prognose der verbleibenden Umläufe bis zur Störung basierte diesmal auf einem einfachen Zustandsmodell, dessen Ergebnisse sie mit der Prognose durch ein Hidden Markov Modell verglichen. Die Zustandstransition, also die Verschlechterung von einem Zustand zum nächsten, wurde stochastisch modelliert. Die entsprechenden Parameter des einfachen Zustandsmodells und des Hidden Markov Modells wurden mit Hilfe der Messdaten identifiziert. Die Evaluation der Ergebnisse erfolgte auch bei diesem Ansatz anhand des Vergleichs der prognostizierten mit den tatsächlich verbleibenden Umläufen bis zur Störung. Erneut steigerten sie die Genauigkeit bei den zehn Vergleichsweichen und auch die Abweichungen zu Beginn wurden kleiner, ein bis vier Umläufe von maximal 20. Leider versäumten es die Autoren auch in diesem Artikel, mehr als 20 Umläufe zu betrachten. Die ohnehin schon fragwürdige Realitätstreue des Experiments (alle ungeschmierten Gleitstühle wurden nach und nach geschmiert, um die Störentstehung rückwärtig zu simulieren), wird durch eine weitere Annahme der Autoren noch zweifelhafter. Ihrem Modell legten sie zugrunde, dass sich alle Weichen nach einer Wartung immer im perfekten Zustand befinden, was zumindest für die DB Netz AG nicht zutrifft (vgl. [GB11, Böh13c]).

Eine reale Anwendung der drei Ansätze, welche in Yilboga et al. [YEGC10], Guclu et al. [GYE⁺10] und Eker et al. [ECG⁺11] vorgestellt wurden, wird skeptisch beurteilt. Obwohl es, vom methodischen Standpunkt aus gesehen, gute Arbeiten sind, werden die Annahmen und experimentellen Bedingungen als zu realitätsfremd angesehen. Mangelnde Schmierung der Gleitstühle ist nur eine von vielen Ursachen für Weichenstörungen. Sie tritt darüber hinaus auch nicht in Erscheinung, indem sukzessive einzelne Gleitstühle ihre Schmierung vollständig verlieren. Eine Prognose von 20 Umläufen – oder weniger, wenn es verlässlich sein soll – ist nicht ausreichend. Gerade stark befahrene, betriebswichtige Weichen laufen innerhalb einer Stunde häufiger um.

Ein französisches Forscherteam veröffentlichte 2011 in Chamroukhi et al. [CSAA11] eine Methode zur Beurteilung eines Weichenantriebs. Die Grundlage bildeten Messungen der elektrischen Antriebsleistung bei 120 Umläufen, welche an einem oder mehreren elektromechanischen Antrieben des Typs VCC aufgezeichnet wurden. Diese Messungen stammen aus dem realen Betrieb der Weiche(n) bei der französischen Bahn SNCF. Die Wissenschaftler bildeten jeden Umlauf mit einem dynamischen auto-regressiven Modell ab, wodurch sie auch das nicht-stationäre Verhalten der Messsignale einbanden. Sie nutzten dieses Modell, um die charakteristischen Messkurven von drei Zuständen (funktionierend, leicht defekt bzw. fehlerhaft oder kritisch defekt bzw. gestört) zu ermitteln. Der Wechsel zwischen den verschiedenen Ausprägungen des dynamischen auto-regressiven Modells, also die Zustandsentwicklung, wurde durch ein Hidden-Logistic-Process Modell abgebildet. Dadurch konnte, auf Basis der Umlaufmessung zum Zeitpunkt t , eine Prognose des Zustands zum Zeitpunkt $t + 1$ berechnet werden. In der gezeigten Evaluation ihrer Methode prognostizierten Chamroukhi et al. dann sukzessive für jeden Zeitpunkt t den kom-

menden Zustand in $t + 1$. Im Vergleich zum tatsächlichen Zustand erreichten sie eine 89.08 Prozent richtige Klassifizierung. Die Prognose, ob der kommende Umlauf fehlerfrei oder gestört sein wird, gibt jedoch nicht ausreichend Zeit um eine Störung zu verhindern oder eine Instandhaltungsmaßnahme gezielt zu planen. Und da eine Anwendung über mehrere Umläufe hinaus ($t + n$ mit $n > 1$) nicht evaluiert oder erwähnt wurde, lässt sich nur vermuten, ob das Modell dabei ebenso gute Ergebnisse liefern würde. Es ist aber davon auszugehen, dass die Genauigkeit sinkt, weil der Zustandsübergang mit dem Hidden-Logistic-Prozess ein stochastisches Modell darstellt. Je weiter damit vorausgerechnet wird, desto größer ist der Korridor, in dem Werte liegen können und desto höher ist die Wahrscheinlichkeit einer falschen Klassifikation. Hinzu kommt, dass die beschriebenen drei Zustände nicht für die Bestimmung eines Abnutzungsvorrates oder einer RUL geeignet sind. Somit ist auch dieses Modell nicht als Zustandsprognose im Sinne der Definition in dieser Arbeit nutzbar.

In der 2011 veröffentlichten Dissertation von Quiroga [Qui11] wurde ein ganzheitlicher Ansatz für die Optimierung des Instandhaltungsprozesses der Gleisgeometrie, genauer des Längshöhenfehlers, vorgestellt. Neben einem Alterungs- und Instandsetzungsmodell sowie einem Optimierungsalgorithmus bildete ein Prognoseverfahren einen Schwerpunkt der Arbeit. Die Alterung, also ein zunehmender Längshöhenfehler wurde mittels eines stochastischen, exponentiellen Modells beschrieben. Dieses Modell wurde anhand von Messdaten aus 20 Jahren einer 200 km langen französischen Hochgeschwindigkeitsstrecke parametrisiert und berücksichtigte auch die Anzahl bisher durchgeführter Wartungen (Stopfungen). Den Zustand prädizierte ein hybrides Modell, dass exponentielle Glättung zweiter Ordnung einsetzte und dabei in drei verschiedenen Systemzuständen (Initial nach Grunderneuerung, Setzung kurz nach Stopfung, Normal) einen unterschiedlich Algorithmus verwendete. Allgemein leistete die Arbeit einen wichtigen Beitrag für die zustandsorientierte Instandhaltung der Infrastruktur. Allerdings wurde nicht auf Weichen und ihr deutlich komplexeres Alterungs- und Störungsverhalten eingegangen. Es findet sich zwar der Hinweis auf speziellen Stopfmaschinen [Qui11, S. 67], die wegen der besonderen Anforderungen bei Weichen notwendig sind, aber in der Modellierung der Alterung oder der Prognose wurde nicht weiter darauf eingegangen. Obendrein ist der untersuchte Zeithorizont von mehreren Monaten und Jahren viel zu grob für Ausfälle bei Weichen. Die Arbeit von Quiroga bietet daher keine Anknüpfungspunkte für Prognose von Weichenstörungen.

Eine relativ neue Arbeit stammt von Asada et al. [ARK13]. Der 2013 erschienene Artikel beschäftigte sich zwar im Kern mit der Zustandsdiagnose von Weichen, enthielt aber auch einen Ansatz für die Prognose. Grundlage bildeten 122 Messungen des Stroms und der Spannung am Motor sowie der Kraft an der Schieberstange. Aufgezeichnet wurden diese in Japan, an einem elektromechanischen Antrieb vom Typ NTS. Dabei wurden die fünf Zustände (fehlerfrei, geringfügig zu hohe und geringfügig zu niedrige Stellkraft sowie deutlich zu hohe und zu niedrige Stellkraft) simuliert. Die Umlaufkurven wurden zur Datenreduktion mit diskreter Wavelet-Transformation bearbeitet und die Klassifizierung der Zustände anschließend mit Stützvektormaschine (engl. Support Vector Machine) (SVM) gelernt. Dabei erzielten die Autoren sehr gute Ergebnisse. Darauf aufbauend versuchten sie Messungen zu klassifizieren, die nicht eindeutig einem der fünf vordefinierten Zustände zuzuordnen waren. Diese Zwischenzustände betrachteten sie als Indikator für die Stärke eines Fehlers und interpretierten dies als Ansatz für eine Prognose, ohne jedoch näher darauf eingegangen zu sein. Da damit aber keine Aussage zum Abnutzungsvorrat oder zur RUL verbunden sind, ist die Nutzbarkeit als Prognosemodell nicht gegeben.

In seiner ebenfalls 2013 veröffentlichten Dissertation griff Asada die Ansätze wieder auf [Asa13]. Erneut bildeten Messungen der Stromstärke und der Spannung am Motor sowie der Kraft an der Schieberstange die Datengrundlage. Wieder wurden die gleichen fünf Zustände simuliert, diesmal jedoch aufgezeichnet an je einem elektromechanischen Antrieb des Typs HW und des Typs Sure-lock in Großbritannien. Um die Stellkraft des kommenden Umlaufs anhand der elektrischen Wirkleistung des aktuellen Umlaufs vorherzusagen, reduzierte Asada die Daten ebenfalls mittels diskreter Wavelet-Transformation, benutzte aber anschließend ein künstliches Neuronales Netz (engl. artificial Neural Network) (aNN) um die Zuordnung zu erlernen und zu testen. Die Prädiktion der Stellkraft für den Zeitpunkt $t + 1$ auf Basis der elektrischen Signale zum Zeitpunkt t erzielte gute Ergebnisse, mindestens 80 Prozent Übereinstimmung zwischen prognostizierter und tatsächlicher Stellkraft. Allerdings empfahl der Autor die Allgemeingültigkeit mit weiteren Daten zu validieren [Asa13, S. 135]. Leider fehlt der Arbeit, neben einer Prognose über den Zeitpunkt $t + 1$ hinaus, auch eine Aussage zum Zusammenhang zwischen Stellkraft und RUL oder Abnutzungsvorrat. Dadurch sind die Ansätze nicht per se für die Weichenzustandsprognose nutzbar.

Quelle	Jahr und Land	Verfahren	Intention	Messgröße	Objekt	Quantität	Analoge Objekteigenschaften bei DB AG	Reale Messdaten verwendet	RUL o. ä. prognostiziert	Prognose mit ausreichender Eingreifzeit
Pedregal et al. [PGMS04]	2004 GB	bivariates Komponentenmodell des Zustandsraumes	Diagnose, aber Prognose der Stellkraft im aktuellen Umlauf	I, F	unbekannter Antrieb	1	O	—	—	—
Pedregal et al. [PGMR09]	2009 GB	bivariates Komponentenmodell des Zustandsraumes	Diagnose, aber Prognose der Stellkraft im aktuellen Umlauf	I, F, Öldruck, Ölstand	elektromechanischer M63 Antrieb	3	—	+	—	—
					elektrohydraulischer Antrieb	4				
Silmon [Sil09]	2009 GB	lineare Regression	Prognose der Fehlerentwicklung	I, F	elektromechanischer M63 Antrieb	3	—	—	—	—
Chattopadhyay und Kumar [CK09]	2009 SE	stochastisches Modell mit Weibull-Verteilung	Prognose der Fehlerentwicklung	kumulierte MGT	Schienen	2	—	+	+	+
Zwannenburger [Zwa09]	2009 CH	lineare Regression	Prognose des Ersatzzeitpunktes	kumulierte Belastung, Liegedauer	Weiche	1736	+	+	+	+
Holzfeind und Hummitsch [HH10]	2010 AT	lineare Regression	Prognose des Ersatzzeitpunktes	Qualitätsindex	Schienen	4000 km	+	+	+	+
García Márquez et al. [GMPR10]	2010 GB	ARMA	Prognose der Umlaufzeit beim nächsten Umlauf	I, F, Öldruck, Ölstand	elektromechanischer M63 Antrieb	3	—	+	—	—
					elektrohydraulischer Antrieb	4				
Gutsche [Gut10]	2010 DE	lineare Regression	Prognose des Ausfalls	P, Umstellvorgänge	elektromechanischer S700K Antrieb	1	+	+	+	+
Yilboga et al. [YEGC10]	2010 TR	aNN	Prognose des Ausfalls durch mangelnde Gleitstuhlschmierung	I, U, F, Stellweg	unbekannter elektromechanischer Antrieb	10	O	—	+	—
Legende: – nein, + ja, o keine Aussage möglich										

Tabelle 2-2: Zusammenfassung bestehender Prädiktionsmodelle für Weichen und Bewertung ihrer Anwendbarkeit

Tabelle 2-2: ... Fortsetzung

Quelle	Jahr und Land	Verfahren	Intention	Messgröße	Objekt	Quantität	Analoge Eigenschaften bei DB AG	Reale Messdaten	RUL o. ä. prognostiziert	ausreichend Eingreifzeit
Gudlu et al. [GYE+10]	2010 TR	ARMA			unbekannter elektromechanischer Gleichstromantrieb	10	—	—	+	—
Eker et al. [ECG+11]	2011 TR	Zustandsmodell mit stochastischen Zustandswechseln, Hidden Markov Modell				10	—	—	+	—
Chamroukhi et al. [CSAA11]	2011 FR	dynamisches autoregressives Modell mit Hidden-Logistic-Process	Prognose des Zustands des nächsten Umlaufs	P	elektromechanischer VCC Antrieb	unbekannt	—	+	—	—
Quiroga [Qui11]	2011 DE	stochastisches Alterungsmodell, Prognose via exponentieller Glättung	Prognose der Entwicklung des Längshöhenfehlers	Längshöhenfehler mm	Schotterbett	200 km	+	+	+	+
Asada et al. [ARK13]	2013 JP	SVM	Diagnose, aber nicht eindeutig klassifizierbare Zwischenzustände als Fehlerstärke für Prognose interpretiert	I, U, F	elektromechanischer NTS Antrieb	1	+	—	—	—
Asada [Asa13]	2013 GB	aNN	Prognose der Stellkraft des nächsten Umlaufs	I, U, F	elektromechanischer HW Antrieb	1	—	—	—	—
					elektromechanischer Sure-lock Antrieb	1				
vorliegende Arbeit zum Vergleich	2017 DE	kNN, aNN, DTree, SVM	Prognose des Ausfalls	P, Bausart, Temperatur, Luftfeuchtigkeit	elektromechanischer S700K Antrieb	29	+	+	+	+
Legende: — nein, + ja, o keine Aussage möglich										

Tabelle 2-2: Zusammenfassung bestehender Prädiktionsmodelle für Weichen und Bewertung ihrer Anwendbarkeit

Tabelle 2-2 fasst die diskutierten Ansätze sowie ihre Bewertung zusammen und führt die vorliegende Arbeit zum Vergleich auf. Aus den genannten Quellen wird ersichtlich, dass bereits Ideen für die Prognose des Weichenzustands erarbeitet wurden, teilweise in Kooperation mit EIU. Allerdings ermöglicht kaum eine Lösung eine Vorhersage im definierten Sinne dieser Arbeit und zwar aus verschiedenen Gründen. Oftmals erfolgte gar keine Prognose oder zumindest keine, die über den nächsten Umlauf hinausging. Wenn doch, war häufig die Eingreifzeit zu kurz, als dass das Instandhaltungspersonal hätte rechtzeitig reagieren oder gar planen können. Viele Ergebnisse beruhten nicht auf realen Daten. Bei den simulierten Daten kam hinzu, dass handelsübliche Sensoren verwendet wurden, die weder robust, noch für Bahnanwendungen zugelassen sind. In vielen Fällen war mit dem Anbringen der Sensoren auch ein Eingriff verbunden, der für den Regelbetrieb eine offizielle Zulassung durch die Aufsichtsbehörde erfordern würde. Zudem sind die Eigenschaften der analysierten Antriebe verschieden von jenen im Netz der DB AG. Insbesondere die in Großbritannien und der Türkei verwendeten Antriebe sind anders aufgebaut, als Antriebe in Deutschland. Zum Beispiel laufen britische Weichen in zwei statt vier bis sechs Sekunden um und werden mit Gleichstrommotoren angetrieben, laut dem InnoTrack-Projektbericht [Uni08, S. 16] ein wichtiger Unterschied bei der Zustandsüberwachung von Weichen. Auch in Frankreich herrschen andere Bedingungen als in Deutschland. Bei der französischen Bahn SNCF ist nur ein einzelner Antriebstyp im Einsatz [GB11]. Bei der DB Netz AG sind es hingegen viele unterschiedliche Typen. Daher ist die Übertragbarkeit solcher Ansätze nicht per se möglich. Des Weiteren basierten die Ergebnisse in den meisten Fällen auf zehn oder weniger untersuchten Objekten bzw. Weichenantrieben. Das ist zwar an und für sich kein Argument gegen eine Anwendung, hinterlässt aber Zweifel an der Generalisierungsmöglichkeit. Die drei insgesamt positiv bewerteten Prädiktionsmodelle sind ausgerechnet jene, die entweder Weichen [HH10] oder den Weichenantrieb [Zwa09] komplett ausklammerten. Oder es sind Ansätze, von denen die Autorin selbst sagt, dass die Anzahl betrachteter Objekte zu klein war [Gut10].

Es ist festzustellen, dass die Anwendbarkeit der diskutierten, bestehenden Prädiktionsmodelle nicht gegeben ist. Es zwingt sich daher die Schlussfolgerung auf, dass die Entwicklung eines eigenen, neuartigen Ansatzes mindestens so erfolgversprechend ist, wie die Weiterführung bestehender Ansätze. Ersteres ist Kern der vorliegenden Arbeit und wird in den folgenden Kapiteln beschrieben.

2.5 Zusammenfassung des Kapitels

In diesem Kapitel wurden die wesentlichen Grundlagen erläutert, um die Zustandsprognose von Eisenbahnweichen in einen breiteren Kontext zu stellen. Für den technischen Zusammenhang wurde die Weiche zunächst in das Gesamtsystem Eisenbahn eingeordnet, gefolgt von einer ausführlichen Beschreibung des Aufbaus und der Funktion einer Weiche. Im Verlauf der Arbeit können so die beschriebenen Interpretationen und Schlussfolgerungen aus den Analysen besser nachvollzogen werden. Die Ausführungen zur Weiche zeigen, wie komplex dieses Infrastrukturelement ist. Eine Komplexität, die die Weiche zu einem kritischen Element macht, gerade vor dem Hintergrund ihrer betrieblichen Relevanz.

Des Weiteren wurde auf die Instandhaltung von Investitionsgütern eingegangen, indem die wichtigsten Strategien erläutert wurden. Es wurde gezeigt, wie die Instandhaltung und ihre Kosten mit den Lebenszykluskosten (engl. Life Cycle Costs) (LCC) in Bezug stehen und die Wirt-

schaftlichkeit beeinflussen. Die zustandsorientierte Instandhaltung ist den anderen Strategien überlegen, für ihre Umsetzung ist aber die Kenntnis des aktuellen und des zukünftigen Zustands Voraussetzung. Im wirtschaftlichen Kontext gesehen, ist die Zustandsprognose also Teil eines ökonomisch effizienten Betriebs bei hoher Verfügbarkeit. Die vorliegende Arbeit erforscht daher die Prognose der verbleibenden Nutzungsdauer, um die Effizienz der Instandhaltung und die Wirtschaftlichkeit des Bahnbetriebs zu erhöhen.

Anschließend wurde die zustandsorientierte Instandhaltung bei der DB Netz AG beschrieben, die dort zwar Ziel, aber nicht Praxis ist. Insbesondere für Weichen fehlt eine Lösung für eine zuverlässige Prognose des Zustands, um Störungen verhindern zu können. Existierende Diagnosesysteme, wie zum Beispiel die vorgestellten Systeme SIDIS W und Roadmaster, bieten dies nicht. Daraus leitet sich auch das Ziel der vorliegenden Arbeit ab: ein Zustandsprognoseverfahren für Eisenbahnweichen. Dabei soll es nicht um die Auswahl einer Instandhaltungsstrategie und deren Ausgestaltung gehen, da im Wesentlichen bereits feststeht, dass eine zustandsorientierte Instandhaltung optimal wäre.

Abschließend wurden als Stand der Forschung bestehende Modelle diskutiert, die eine Prädiktion des Weichenzustands zum Inhalt haben. Unter Berücksichtigung verschiedener Aspekte (z.B. analysierte Objekte, Ähnlichkeiten zu Weichen der DB AG, Verwendung realer Messdaten, etc.) zeigt die Diskussion, dass international mehrere Ansätze zur Lösung vorgeschlagen werden, von linearer Regression, über stochastisch modellierte Zustandsübergänge bis hin zu Stützvektormaschine (engl. Support Vector Machine) (SVM). Allerdings ermöglicht keine der Lösungen eine verlässliche Vorhersage der verbleibenden Nutzungsdauer (RUL) oder einer äquivalenten Größe des Abnutzungsvorrates. Daraus folgt wiederum, dass es unter Einsatz der vorgestellten Ansätze unmöglich bleibt, rechtzeitig Instandhaltungsmaßnahmen einzuleiten bzw. eine zustandsorientierte Instandhaltungsstrategie für Weichen zu realisieren. Es wird deshalb als notwendig erachtet, einen eigenen Ansatz für die Zustandsprognose von Eisenbahnweichen zu entwickeln, um die Forschungsfragen der Arbeit (siehe Abschnitt 1.2) beantwortet zu können.

3 Verfahren der Datenanalyse und der Zustandsprädiktion

Im Mittelpunkt von Kapitel 2 standen die Grundlagen zum Aufbau und zur Funktion von Weichen sowie der Zusammenhang zwischen Lebenszykluskosten (engl. Life Cycle Costs) (LCC), Instandhaltungsstrategie und Zustandsbeurteilung. Dabei wurden wichtige technische und wirtschaftliche Sachverhalte erläutert, die zum Verständnis der Arbeit notwendig sind. Außerdem wurden die zustandsorientierte Instandhaltung bei der Eisenbahninfrastruktur und Überwachungssysteme für Weichen beleuchtet. Daraus leitete sich ab, dass die Prognose erforscht werden muss, um mit einer zustandsorientierten Instandhaltung Kosten zu verringern und die Pünktlichkeit zu erhöhen. Im aktuellen Kapitel geht es um Grundlagen zu Verfahren der Datenanalyse und der Zustandsprädiktion als Stand der Forschung. Zunächst werden allgemeine Ansätze zur Zustandsbeurteilung verglichen (Abschnitt 3.1), um eine fundierte Herangehensweise zu identifizieren, an der sich diese Arbeit orientieren kann. Danach werden einzelne Verfahrensgruppen kurz vorgestellt und bewertet, die für eine Prädiktion infrage kommen (Abschnitt 3.2). Es folgt eine ausführliche Diskussion und Bewertung von Metriken, mit denen sich die Qualität von Prognosen beurteilen lässt (Abschnitt 3.3). Dies ist insbesondere wichtig, weil die Wahl einer passenden Metrik für die Wahl eines guten Prognoseverfahrens entscheidend ist. Auf die Bewertungen dieses Kapitels wird im Laufe der Arbeit zurückgegriffen, wenn Prognoseverfahren und Metriken auszuwählen sind, die sich für die vorliegende Problem- und Datenstruktur eignen.

3.1 Allgemeine Ansätze für die Zustandsbeurteilung

Ob bei Kraftwerksanlagen, Öltraffinerien oder Flugzeugen, überall dort wo die Vorteile der zustandsorientierten Instandhaltung genutzt werden sollen, sind ein Anlagenmonitoring sowie Diagnose und Prognose unverzichtbar (vgl. Abschnitt 2.2.4 und 2.2.5). Aus diesem Grund hat sich in der letzten Dekade ein Zweig der Ingenieurwissenschaften herausgebildet, der sich explizit mit dem Thema befasst [UGL08]. Bekannt als Prognose und Zustandsmanagement (engl. Prognostics and Health Management) (PHM), steht dabei die Verwertung von Messungen, Modellen und Software für die Fehlerdetektion und -vorhersage im Fokus [KBR06, S. 354]. PHM verfolgt das Ziel, dem Anwender eine Entscheidungsunterstützung zu bieten. Im Folgenden werden verschiedene Ansätze des PHM vorgestellt, ihre Eignung für die Weichenzustandsprognose diskutiert und anschließend mit der Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD) als Alternative verglichen.

3.1.1 Ansätze des Prognose- und Zustandsmanagements

Im Wesentlichen finden beim PHM drei grundsätzlich verschiedene Ansätze Anwendung, um zu Aussagen über die verbleibende Nutzungsdauer (engl. Remaining Useful Life) (RUL) zu gelangen (vgl. [SKW08, UGL08]).

- Die **zuverlässigkeitsbasierten Methoden** bilden die Gruppe der einfachsten Methoden. Sie basieren auf einer statistisch ermittelten Wahrscheinlichkeit, dass eine Anlage innerhalb einer definierten Zeitspanne funktioniert [SKW08]. Bei der Zuverlässigkeitstheorie werden

die zugrundeliegende Wahrscheinlichkeitsverteilung und ihre Parameter meist in der Produktentwicklung ermittelt. Dabei werden mittels simulierter beschleunigter Alterung oder Belastung Kenngrößen wie die mittlere Zeit zwischen einem Ausfall (engl. Mean Time Between Failure) (MTBF) oder die mittlere Zeit bis zu einem Ausfall (engl. Mean Time To Failure) (MTTF) bestimmt. Außerdem wird eine Verteilungsfunktion geschätzt, bspw. die Exponentialverteilung oder die Weibull-Verteilung [Fis84]. Die Schwäche der zuverlässigkeitsbasierten Methoden ist die starke Verallgemeinerung, die kaum individuelle Einsatz- und Umgebungsbedingungen einer Anlage berücksichtigt [UGL08, BBKJ08]. Daher wird dieser Ansatz nur selten für die operative Zustandsbeurteilung angewendet, vor allem nicht, wenn es sich um eine komplexe Anlage wie die Weiche handelt.

- Eine weitere Gruppe bilden die **modellgetriebenen Methoden**. Als Modell wird die Abbildung der physikalischen Gesetzmäßigkeiten bezeichnet, die zur Abnutzung führen. Wenn richtig erstellt, gelten sie als der genaueste Ansatz für die Vorhersage der Restlebensdauer. Besonders erfolgreich sind modellgetriebene Ansätze bei Phänomenen, die auf Materialalterung beruhen, z.B. das Wachstum eines Schienenrisses. Die hohe Genauigkeit setzt allerdings voraus, dass die physikalischen Vorgänge bekannt sind und modelliert werden können, was meist auch mit einem hohem Rechenaufwand einhergeht. Außerdem ist die Pflege der Modelle aufwändig, insbesondere wenn neue Anlagenvarianten hinzukommen. So erklärt sich auch die verhältnismäßig geringe Zahl an modellgetriebenen Ansätzen bei komplexen mechanischen und elektronischen Anlagen. [SKW08, UGL08, GKVJ13]
- Die dritte Gruppe von Ansätzen bilden die **datengetriebenen Methoden**. Die Grundlage bilden empirische Daten aus dem Betrieb und der Überwachung einer Anlage. Es werden charakteristische Muster in diesen Daten identifiziert, anhand derer die RUL der Anlage bestimmbar ist. Für diesen Ansatz ist das detaillierte Verständnis von Physik weniger bedeutsam, und dennoch können die individuellen Umgebungsbedingungen einer Anlage berücksichtigt werden. [SKW08] Zudem kann die Toolkette für den datengetriebenen Ansatz automatisiert werden, wodurch der Pflegeaufwand gering bleibt. Neben der Bedingung, dass überhaupt empirische Daten vorliegen müssen, stellt die Datenqualität die entscheidende Hürde für eine erfolgreiche Anwendung dar. Es muss mit mehrdimensionalen, unvollständigen und verrauschten Daten umgegangen werden, von denen oft nicht bekannt ist, inwieweit sie die inneren Abnutzungsphänomene spiegeln. [UGL08]

Im Fall von Eisenbahnweichen scheiden zuverlässigkeitsbasierte Methoden für die Zustandsprognose aus, weil die Umgebungsbedingungen, die Belastung und die Bauart zu individuell sind. Von den verbleibenden Ansätzen verspricht ein modellbasiertes Vorgehen zwar die höchste Genauigkeit, aber bei der Komplexität des Gesamtsystems Weiche mit seinen vielen mechanischen, elektromechanischen und elektronischen Bauteilen (vgl. Abschnitt 2.1.2) wird die Modellerstellung sehr zeitaufwendig, wenn sie überhaupt gelingt. Die von der Studie EcoSwitch identifizierten 75 Einflussfaktoren auf den Zustand (vgl. [JZ02]) zu modellieren, erscheint zudem weniger aussichtsreich als einen datengetriebenen Ansatz anzuwenden [Böh13d]. Diese Einschätzung wird dadurch bekräftigt, dass sich in der Luftfahrt der datengetriebene Ansatz bewährt hat und deshalb generell empfohlen wird, um komplexe Anlagen zu überwachen (vgl. [Hes02]).

3.1.2 Datengetriebene Ansätze des Prognose- und Zustandsmanagements im Vergleich zur Wissensentdeckung in Datenbanken

Um ein qualitativ hochwertiges Ergebnis sicherzustellen, soll sich die Entwicklung des Prognosemodells bestmöglich an einem systematischen Prozess orientieren. Der vorherige Abschnitt diskutierte deshalb die Ansätze des PHM und stellte heraus, dass sich ein datengetriebener Ansatz am ehesten eignen würde. In diesem Abschnitt wird eine Alternative erörtert. Bei näherer Betrachtung des datengetriebenen PHM werden das Ziel, das Vorgehen, die Methoden und die Ergebnisevaluation von einer etwas älteren Wissenschaftsdisziplin eingeschlossen oder überschneiden sich zumindest stark mit dieser. Gemeint ist die Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD).

„KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data“ [FPSS96a, S. 2]

Auch beim KDD steht also der Erkenntnisgewinn zur Entscheidungsunterstützung im Vordergrund. Dazu werden große, mehrdimensionale, empirische Datenbestände analysiert. Doch wo PHM das relevante Wissen auf die Restlebensdauer beschränkt, ist KDD weiter gefasst. KDD sucht explizit auch nach bisher unbekannten Phänomenen und Zusammenhängen.

Ein Vergleich des KDD-Prozesses, wie er von Fayyad et al. [FPSS96b] definiert wurde, mit dem Prozess des Open System Architecture Condition Based Management Standard beim PHM, wie in Sheppard et al. [SKW08] beschrieben, verdeutlicht sowohl die Unterschiede als auch die Gemeinsamkeiten (siehe Abbildung 3-1). Beim KDD bilden unterschiedlichste Datenbestände die Grundlage, während beim PHM Messungen und Sensoraufzeichnungen die Ausgangsbasis sind. Im KDD-Prozessschritt der Datenselektion werden diese unterschiedlichen Datenbestände verknüpft, wodurch das einbezogene implizite Wissen vergrößert wird.

Des Weiteren behandelt der Prozess der KDD die Datenselektion, die Datenvorverarbeitung und die Datentransformation als eigenständige und ebenso wichtige Schritte wie die Mustersuche (Data Mining), Interpretation und die Evaluation. Diese essentiellen Schritte werden im PHM nicht ausreichend berücksichtigt. Wie in Hess und Fila [Hes02, S. 6-2792] nachzulesen ist, scheiterten viele Anwendungen von PHM an ineffektiver Sortierung, Handhabung und Nutzung von Monitoringdaten. Und auch Wheeler et al. [WKP10, S. 16] haben das Datenmanagement als einen Bereich identifiziert, der im PHM nicht ausreichend adressiert wird. Häufig existieren in Unternehmen (elektronische) Aufzeichnungen, die für eine Zustandsbeurteilung relevant sein können. Hess und Fila [Hes02] sowie Böhm und Gutsche [BG11] plädieren dafür, solche Daten zu nutzen, anstatt Sensoren zur Erfassung äquivalenter Größen anzuschaffen.

Als Data Mining wird der zentrale Schritt der Wissensentdeckung bezeichnet. Unter dem Begriff Data Mining werden systematische, mathematisch fundierte Methoden zur Erkennung von Mustern und Regeln in großen Datenbeständen zusammengefasst (vgl. [FPSS96b, S. 40] und [Pet05, S. 10]). Gemäß der jeweiligen Zielstellung werden von Petersohn [Pet05] die folgenden Anwendungsbereiche unterschieden:

- Die **Klassenbildung** identifiziert in den Daten bisher unbekannte Gruppen, deren Objekte einander ähnlich sind bzw. von den Objekten in anderen Gruppen unterscheiden.

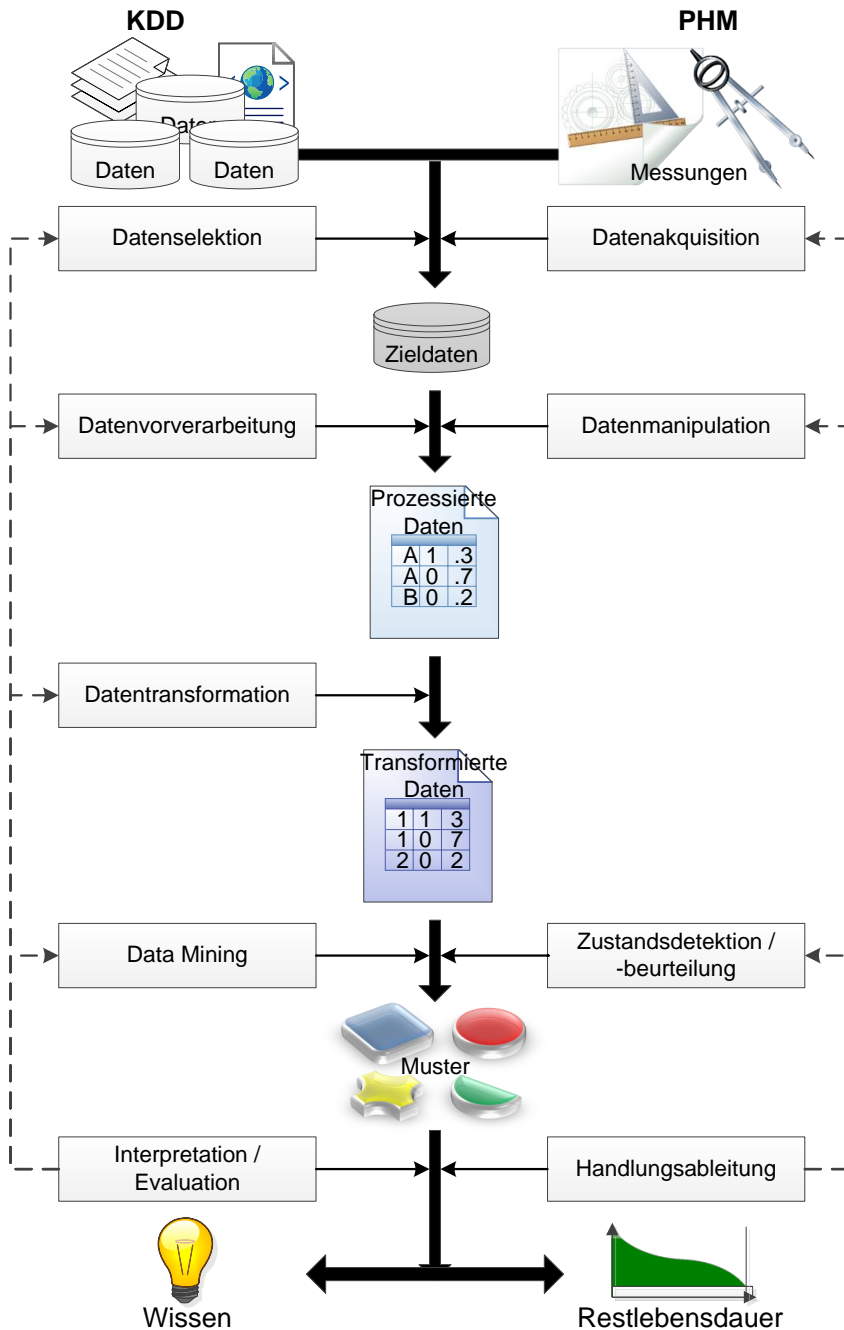


Abbildung 3-1: Vergleich des Prozesses bei der KDD (in Anlehnung an [FPSS96b]) und beim PHM (in Anlehnung an [SKW08])

- Sollen Regeln für Ereignisse gefunden werden, die in Abhängigkeit voneinander auftreten, wird dies als **Assoziation** bezeichnet.
- Bei der **Klassifizierung** wird das Wissen über die Zuordnung von Objekten zu bekannten Gruppen extrahiert.
- Die **Zeitreihenanalyse** modelliert funktionale zeitliche Zusammenhänge, so dass für neue Argumente der Funktionswert bestimmt werden kann.

Es existiert eine große Anzahl von konkreten Verfahren und Techniken, die im Data Mining angewendet werden. Eine Untermenge derselben Verfahren, hauptsächlich aus den Anwendungsbereichen der Klassifikation und der Zeitreihenanalyse, kommt auch bei Zustandsdetektion und -beurteilung im PHM zum Einsatz. Sehr selten werden beim PHM Verfahren zur Klassenbildung und zur Assoziation genutzt, obwohl diese beim Data Mining gerade die bisher unbekannten Zusammenhänge aufdecken.

Der abschließende Prozessschritt der Wissensentdeckung ist die Interpretation und Evaluation der gefundenen Muster. Da die konkreten Verfahren und Techniken der Zustandsbeurteilung eine Untermenge des Data Minings sind, sollten auch die Verfahren der Evaluation die gleichen sein. Aber eine ausreichende Diskussion zu Evaluationsverfahren und Performance Metriken findet beim PHM nicht statt bzw. beginnt gerade erst (vgl. [SCB⁺08, LYRF08, SCS⁺10, WKP10, LGY11]). Demgegenüber wird beim Data Mining seit längerem eine intensive Debatte um die Evaluation und Metriken geführt. Insbesondere in Bezug auf die Zielerreichung der zustandsorientierten Instandhaltung (Reduktion der Instandhaltungskosten), ist die Ergebnisevaluation unablässig. Sie wird daher im Kapitel 4 ausführlich untersucht.

Unter der Voraussetzung, dass ein datengetriebener Ansatz für die Zustandsprognose von Eisenbahnweichen verfolgt wird, ist eine Ausrichtung am KDD unter Anwendung von Data Mining besser geeignet, als eine Orientierung am PHM, da:

1. eine sehr umfangreiche Datenbasis zu erwarten ist (bspw. ergäben Überwachungsdaten aus einem Jahr von 10 Weichen, die einmal stündlich umlaufen, bereits 87600 Datensätze. Das Netz der DB AG umfasst allerdings 67071 Weichen und Kreuzungen [DB 16], von denen viele mehrmals in der Stunde umlaufen.),
2. die vielen Einflussfaktoren auf den Weichenzustand aus unterschiedlichen Quellen für die Analyse zusammengeführt werden müssen (z.B. Instandhaltungsdokumentation, Klimadaten, Messdaten),
3. auch versteckte Zusammenhänge und unbekannte Gruppen zu identifizieren sind, wie bspw. Weichen mit ähnlichem Abnutzungsverhalten,
4. die Evaluation der konkreten Prognoseverfahren und der Zielerreichung der Zustandsprognose sehr wichtig sind.

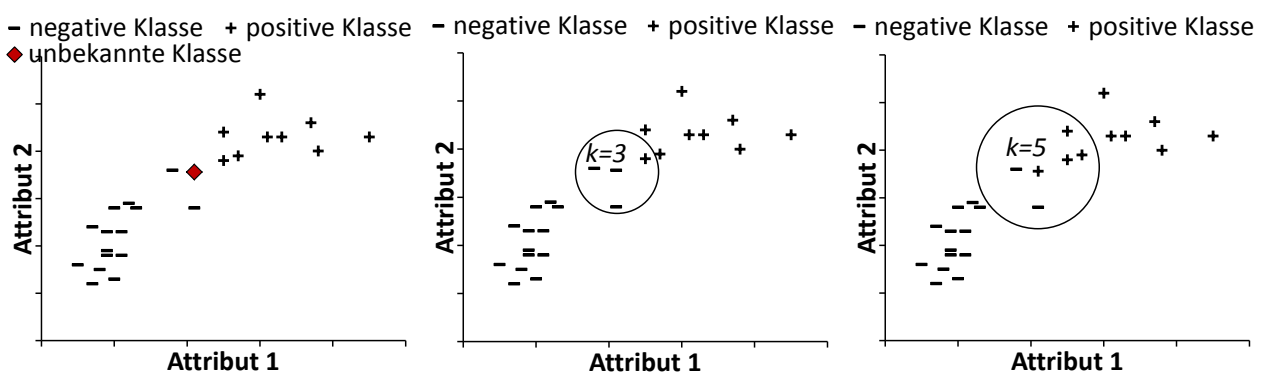
Und noch ein weiterer Punkt spricht für Data Mining – auch wenn dieser eine unbeabsichtigte Folge der Wortwahl ist. Hess und Fila kommen selbst zu dem Schluss, dass vorhandene Daten besser genutzt werden sollen: *„Efforts need to begin that will mine legacy data for failure information ...“* [Hes02, S. 6-2796]

3.2 Methoden für Prädiktionsmodelle

Im aktuellen Kapitel wurde im Abschnitt 3.1 mit dem KDD ein allgemeiner Prozess ausgewählt, an dem sich die Entwicklung der Zustandsprognose orientieren soll. Für die Zustandsprognose stehen zahlreiche Verfahren und Techniken zur Auswahl. Ganze Lehrbücher widmen sich ihrer Erläuterung und der Diskussion ihrer Vor- und Nachteile (z.B. von Hand et al. [HMS01], Petersohn [Pet05], Bramer [Bra07], Maimon und Rokach [MR10] oder Linoff und Berry [LB11]). Und nach wie vor entwickelt die Wissenschaft Verbesserungen und Neuerungen. Das sogenannte *No Free Lunch Theorem*, welches von Duda et al. 1973 eingeführt wurde, besagt im Wesentlichen, dass es kein ultimativ bestes Verfahren zur Mustererkennung gibt [DHS12]. Vielmehr kommt es auf die zugrundeliegende Problemstellung und die Struktur der Ausgangsdaten an [FPSS96b, S. 48]. Es ist daher weder sinnvoll noch möglich, im Rahmen dieser Arbeit alle Verfahren detailliert zu erläutern. Um dennoch einen Einblick zu geben, wird an dieser Stelle eine Auswahl von häufig verwendeten Verfahren bzw. Verfahrensguppen vorgestellt. Die jeweiligen Eigenschaften werden bewertet, damit später ein oder mehrere geeignete Verfahren ausgewählt werden können.

3.2.1 k-Nearest-Neighbor

Zu den einfachsten Verfahren zählt die Klassifikation auf Basis der k -nächsten-Nachbarn (engl. k -Nearest-Neighbor) (kNN). Wobei k eine beliebige, natürliche Zahl repräsentiert, die vom Anwender festgelegt wird. Dahinter steht die Annahme, dass ein neuer Datenpunkt derselben Klasse angehört, wie Datenpunkte, die ihm sehr ähnlich sind. Soll ein neuer Datenpunkt einer Klasse zugeordnet werden, müssen zunächst die k nächsten Nachbarn im Eigenschaftsraum gefunden werden. Die Nähe wird durch ein Proximitätsmaß bestimmt, bspw. die Euklidische Distanz zweier Datenpunkte. Dann wird dem neuen Datenpunkt die Klasse zugewiesen, welche am häufigsten unter diesen k Nachbarn vorkommt. Alternativ werden die Nachbarn mit ihrem Abstand gewichtet, so dass die Klassen von ähnlicheren Nachbarn bei der Klassifikationsentscheidung stärker berücksichtigt werden. [HMS01]



(a) Datensätze mit bekannter Klasse (b) Klassifikationsergebnis der drei nächsten Nachbarn (c) Klassifikationsergebnis der fünf nächsten Nachbarn

Abbildung 3-2: Beispiel der Klassifikation eines unbekannten Datensatzes mit kNN

Für kNN ist kein Training notwendig, da kein Modell gelernt werden muss. Ein Algorithmus ist daher leicht zu implementieren und funktioniert unabhängig von der Klassenanzahl. Zudem ist kNN relativ robust gegenüber unvollständigen Daten. Wie am Beispiel in Abbildung 3-2 zu sehen ist, kann allerdings die Klassenzuweisung sehr unterschiedlich ausfallen, je nach Wert

für k . Es obliegt daher dem Anwender, gute Werte für k empirisch zu ermitteln. [HMS01] Bei hoher Anzahl an Attributen, insbesondere wenn davon nur wenige charakteristisch für die Klassenzugehörigkeit sind, liefert kNN nur gute Ergebnisse, wenn auch genügend Datenpunkte mit den Beispielsklassen präsent sind. Die Ergebnisqualität sinkt überproportional mit wachsenden Dimensionen, da das Distanzmaß die alleinige Basis für die Klassenbildung ist. Schon bei 10 bis 20 Dimensionen kann das Ergebnis bedeutungslos werden (vgl. [BGRS99, HAK00]). Neben dieser ausgeprägten Sensitivität gegenüber der Dimensionsanzahl existiert noch ein weiterer entscheidender Nachteil. Für die Klassifikation bzw. die Nachbarschaftssuche muss immer der gesamte Datensatz vorhanden sein, und es müssen auch die Distanzen zu allen Datenpunkten ermittelt werden. Das kann vor allem bei permanent wachsenden Datenbeständen zum Problem bei der Rechenzeit werden.

3.2.2 Bayes'sche Netze

Bayes'sche Netze (engl. Bayesian Networks) bestehen aus einem gerichteten azyklischen Graphen, der die Struktur von einander bedingenden Zufallsgrößen abbildet. Die Knoten stellen die Zufallsgrößen mit ihrer angenommenen Wahrscheinlichkeitsverteilung dar. Die Kanten repräsentieren die bedingte Abhängigkeit. [Bra07] Bayes'sche Netze folgen der Annahme, dass die relative Häufigkeit der Zielklassenwerte und der Attribute repräsentativ für die Problemstellung sind. Dieses A-priori-Wissen wird dazu verwendet, eine bedingte Wahrscheinlichkeit für die Klassenzugehörigkeit zu bestimmen, welche sich durch einen neuen Datensatz mit Attributwerten ergibt. Die Klassifizierung des neuen Datensatzes erfolgt meist anhand der Klasse mit der höchsten Wahrscheinlichkeit.

Wegen ihrer klaren semantischen Darstellung sind Bayes'sche Netze für den Menschen gut verständlich, eignen sich aber wenig für Probleme mit vielen Zielklassen [MR05]. Wenn die Struktur des Netzes bereits bekannt ist, sind Bayes'sche Netze verhältnismäßig einfach anzuwenden und erzielen gute Ergebnisse [Bra07]. Andernfalls kann das Erlernen der Netzstruktur, je nach Datenbasis, sehr komplex und aufwendig werden [Lee10]. Außerdem ist für Bayes'sche Netze ungeklärt, wie sie mit Klassen umgehen, die aus kategorischen und stetigen Klassenmerkmalen bestehen. Der weitaus schwerwiegendere Nachteil ergibt sich aus der Verwendung der geschätzten relativen Häufigkeit der Klassenvertreter. Je weniger balanciert die Klassen sind, desto eher tendiert das Netz dazu, für neue Daten einen Vertreter zu bestimmen, der einfach der Klasse entstammt, die bisher am häufigsten anzutreffen war. [SAR10]

3.2.3 Künstliche Neuronale Netze

Ein künstliches Neuronales Netz (engl. artificial Neural Network) (aNN) nimmt das Gehirn zum Vorbild. Wie bei dem biologischen Original wird mit Hilfe von einzelnen Neuronen eine Eingangsinformation verarbeitet. Die Summe der gewichteten Signale der Eingangsschicht wird innerhalb des Neurons, in Abhängigkeit vom Schwellwert der Aktivierungsfunktion, zu einem Ausgangssignal weiterverarbeitet. So entsteht ein neues Eingangssignal für weitere Neuronen (siehe Abbildung 3-3). In der Lernphase des Netzes werden die Gewichte in jeder Schicht adaptiert, bis der Gesamtfehler der Ausgangsschicht (oder ein anderes Gütekriterium) ein Mindestmaß unterschreitet. Dadurch können komplexe, nicht lineare Zusammenhänge zwischen einem Vektor von Eingangssignalen und einem Vektor von Ausgangssignalen abgebildet werden. [HMS01]

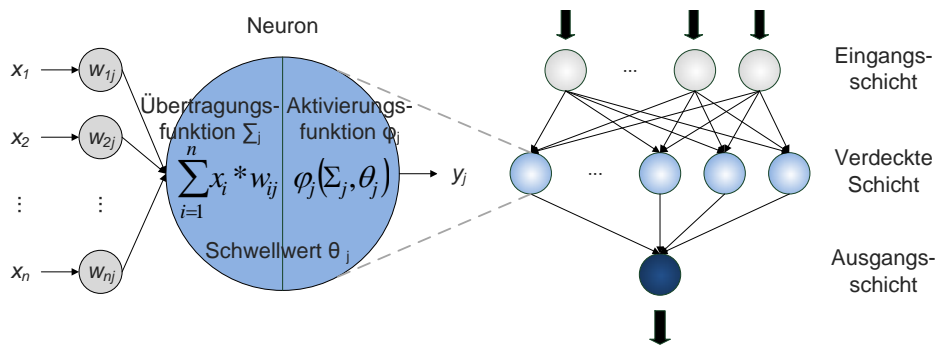


Abbildung 3-3: Beispiel für den Aufbau eines Neurons und der Struktur eines aNN

Die grundlegende Theorie zu aNN entstand schon in den 1940er Jahren. Doch erst in den 1980er Jahren entwickelte die Wissenschaft Netzstrukturen und Trainingsalgorithmen, die zur effizienten Lösung komplexer Problemstellungen geeignet sind. [Zha10] Nicht zuletzt verhalfen Computer den aNN zum Durchbruch, denn die aufwendigen Berechnungen während der Trainingsphase wären ohne elektronische Rechentechnik nicht in angemessener Zeit zu bewältigen. So gibt es mittlerweile vielfältige Netztypen für unterschiedliche Anwendungsfelder, über die Zhang einen guten, einführenden Überblick gibt [Zha10].

Vorteilhaft an aNN ist nicht nur ihre Fähigkeit komplexe, nicht lineare Zusammenhänge nachzubilden zu können. Ihr Design ist bestens geeignet, um die Informationsverarbeitung zu parallelisieren. Außerdem wird zum Training von aNN kein A-priori-Wissen benötigt, wodurch aNN breiter einsetzbar sind. Ein weiterer Vorteil liegt in ihrer Toleranz gegenüber unvollständigen, verrauschten Daten. [Zha10] Diesen Vorteilen stehen auch einige Nachteile gegenüber. So kann die Trainingsphase sehr rechenintensiv sein [Lee10]. Wichtiger ist jedoch, dass aNN dazu tendieren, lokale Optima zu finden. Es ist daher für den Anwender nicht sicher bzw. schwer prüfbar, ob ein gelerntes Netz wirklich das globale Optimum zur Problemlösung repräsentiert. Hinzu kommt ein gewisser Hang zur Überanpassung (engl. Overfitting). Hierbei ermittelt das Netz die Lösung genau passend zum Trainingsdatensatz, ist dafür aber nicht mehr in der Lage zu verallgemeinern. [HMS01] Der Anwender muss dieser Überanpassung aktiv entgegenwirken. Es gibt kein Standardverfahren, um eine passende Netzstruktur auszuwählen [Lee10]. Letztlich muss der Anwender seine Daten, seine Problemstellung und die vielfältigen Möglichkeiten der Netztopologie gut kennen, um einen Nutzen aus dem Einsatz künstlicher Neuronaler Netze zu ziehen [Zha10].

3.2.4 Entscheidungsbaumverfahren

Ein Entscheidungsbaum (engl. Decision Tree) (DTree) ist ein hierarchisch gerichteter Graph, dessen Blätter die Klassen darstellen. Die übrigen Knoten repräsentieren ausgewählte Attribute des Datensatzes. Die ausgehenden Kanten sind jeweils Ausprägungen dieses Attributs, nach denen der Datensatz unterteilt wird. Um einen Entscheidungsbaum zu erstellen, werden alle Trainingsdaten in der Wurzel zusammengekommen. Mit Hilfe eines Teilungskriteriums wird das Attribut ausgewählt, welches zur Bestimmung der Klassen den höchsten Informationsgehalt besitzt. [Pet05] Häufig wird die Entropie oder der Gini Index als Teilungskriterium verwendet, aber auch andere Kriterien sind möglich [RM10]. Anhand des ausgewählten Attributs bzw. anhand dessen Ausprägungen werden Teilmengen in zwei oder mehr Knoten gebildet. Iterativ wird nun in jedem Knoten, analog zur Wurzel, wieder das Attribut mit dem höchsten Informationsgehalt

zur Teilung bestimmt. Von der Wurzel bis zu den Blättern entstehen disjunkte Teilmengen des Datensatzes, die in den Blättern nur noch Datenpunkte einer Klasse enthalten. Abbildung 3-4 veranschaulicht den Prozess und den Aufbau eines Entscheidungsbaumes. In der realen Anwendung entsteht zunächst ein weit verzweigter DTree, aufgrund von widersprüchlichen Daten. Das Resultat ist eine Überanpassung, der aber in den meisten Implementierungen durch das Beschneiden (engl. Pruning) statistisch irrelevanter Zweige entgegengewirkt wird [Pet05].

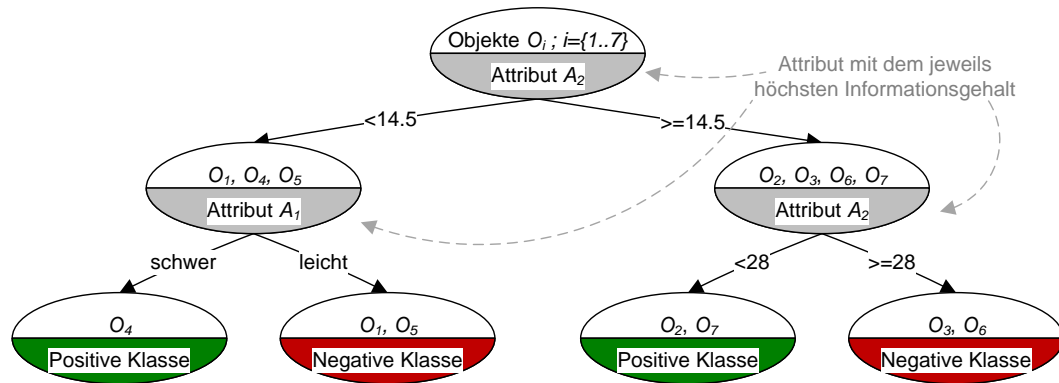


Abbildung 3-4: Beispiel eines einfachen Entscheidungsbaumes

Obwohl schon Ende der 1980er und Anfang der 1990er Jahre entwickelt, gelten Entscheidungsbaumverfahren, insbesondere der C4.5-Algorithmus von Quinlan [Qui93], als ein wichtiges Referenzverfahren innerhalb des Data Mining [Kös08, S. 52]. DTree haben den großen Vorteil, dass sie gleichzeitig mit Attributen von unterschiedlichem Skalenniveau umgehen können. Außerdem verarbeiten sie mühelos unvollständige Daten. Die Trainingsalgorithmen sind zwar nicht so einfach zu implementieren, insbesondere wenn Zweige beschnitten werden sollen. Aber ihre Anwendung kann ohne besonderes Wissen über die Verteilung oder die Klassenstruktur erfolgen. DTree sind selbsterklärend und können von Menschen auch ohne spezielle Expertise gut interpretiert werden. Nachteilig an vielen Standardalgorithmen ist, dass sie diskrete Klassenwerte verlangen. Wegen der Teilung in disjunkte Mengen liefern Entscheidungsbaumverfahren keine guten Ergebnisse, wenn viele, miteinander verbundene Attribute die Klassen beschreiben. [RM10] Außerdem reagieren DTree auf verrauschte Daten mit relativ starker Überanpassung [Qui93].

3.2.5 Support Vector Machines

In den letzten Jahren sind verschiedene Arten der sogenannten Stützvektormaschine (engl. Support Vector Machine) (SVM) zur Klassifikation populär geworden. Das 1995 von Vapnik [Vap08] eingeführte Verfahren löst ein duales Optimierungsproblem. Es zielt – einfach ausgedrückt – darauf ab, eine Hyperebene (Separationsachse) im Eigenschaftsraum zu identifizieren, die einen möglichst großen Abstand (Marge) zu den Datenpunkten aufweist und dabei die Klassen möglichst vollständig voneinander trennt [Bur98]. Definiert wird diese Hyperebene anhand von Datenpunkten entlang der jeweiligen Margengrenze. Sie bilden die Stützvektoren. Die Abbildung 3-5 veranschaulicht das Prinzip für den zweidimensionalen Fall einer binären Klassifikation.

Bei den meisten realen Fragestellungen ist eine lineare Trennung der Klassen unmöglich. Zur Lösung solcher Probleme transformieren SVM den originalen Eigenschaftsraum in einen höherdimensionalen Raum, der eine einfache Separation ermöglicht. Dazu bedienen sie sich einer

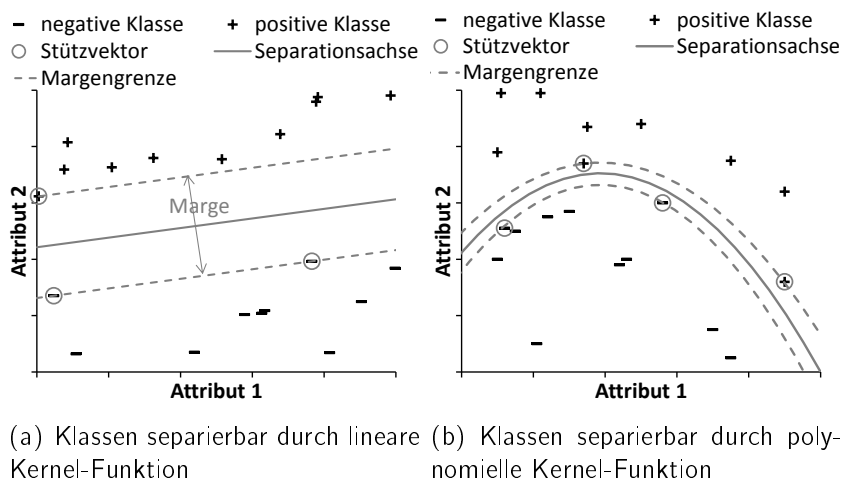


Abbildung 3-5: Beispiel für SVM mit verschiedenen Kernel-Funktionen

Kernel-Funktion, die mit geringem Rechenaufwand hin- und rücktransformiert werden kann und die Hyperebenen zu beschreiben vermag, wie bspw. in Abbildung 3-6. Allerdings eignen sich die verschiedenen Kernel-Funktionen unterschiedlich gut zur Definition der Hyperebene, je nach Datenstruktur. Daher verwundert es nicht, dass zahlreiche, parametrierbare Kernel-Funktionen existieren, z.B. lineare, polynomielle n-ten Grades, sigmoid oder Radial-Base-Function (RBF) Kernels (vgl. [Shm10]).

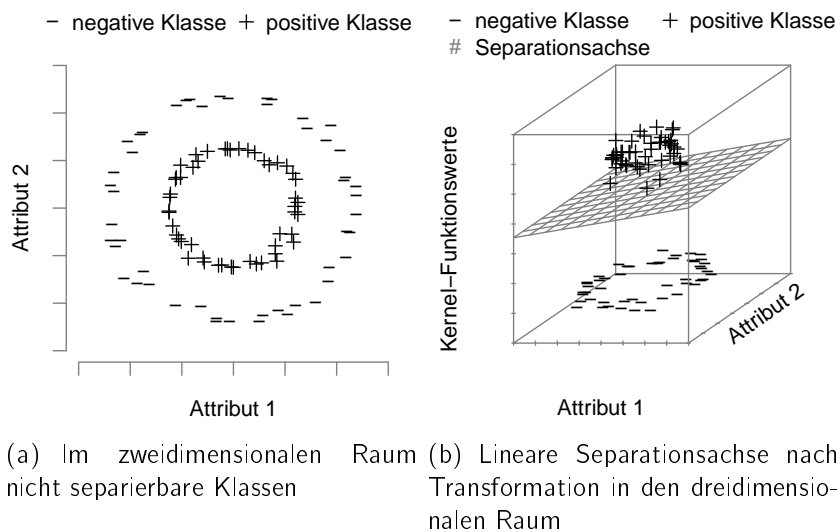


Abbildung 3-6: Beispiel für SVM-Transformation in einen höherdimensionalen Eigenschaftsraum mittels RBF-Kernel

Support Vector Machines brauchen kein vorhergehendes Wissen über die Problemstruktur und können auch mit permutierten, verfälschten Daten umgehen. Sie zeichnen sich durch ein sehr gutes Verallgemeinerungsvermögen aus, wodurch sie relativ unempfindlich für Überanpassung sind. [Bur98] Auch die Anzahl der Dimensionen ist für die Ergebnisqualität weniger relevant als bei vielen anderen Verfahren [BHW10]. Die Lösungen repräsentieren immer das globale Optimum, auch wenn die Stützvektoren für Menschen schwer zu interpretieren sind [Shm10]. Dieser Vorteil entsteht aus der Transformation in den höherdimensionalen Raum. In Verbindung mit der Stützvektorsuche bzw. der dualen Optimierung wird der Vorteil allerdings durch einen extrem hohen Rechenaufwand erkauft. Multi-Klassen-Probleme sind zwar lösbar, erhöhen aber den Re-

chenaufwand der Trainingsphase exponentiell. [BHW10] Zudem sind SVM nicht ohne Weiteres auf nominal- oder ordinalskalierte Daten anwendbar [Bur98]. Die größte Anwendungshürde stellt die Wahl des Kernels dar, da sie für die Ergebnisqualität entscheidend ist. Letztlich muss der Anwender von SVM die verschiedenen Kernel-Funktionen kennen und jene Parameter empirisch ermitteln, für die die besten Klassifikationsergebnisse erzielt werden.

3.2.6 Vergleichende Bewertung der Verfahren

Methode	kNN	Bayes'sche Netze	aNN	DTree	SVM
Eigenschaft					
Implementierungsaufwand	++	+	–	– –	–
Trainingsaufwand	++	+	–	+	– –
Parametrierungsaufwand	++	+	–	++	– –
Speicherbedarf	– –	++	+	+	+
Rechenaufwand zur Klassifikation	– –	++	++	–	–
Sensitivität gegenüber Klassenanzahl	+	– –	+	+	+
Abhängigkeit von A-priori-Wissen	++	– –	++	++	++
Eignung bei unvollständigen, verrauschten Daten	+	– –	+	+	++
Robustheit gegenüber unbalancierten Klassen	+	– –	+	+	++
Eignung bei mehr als 20 Dimensionen	– –	– –	+	–	++
Fähigkeit zur Bestimmung des globalen Optimums	+	–	– –	–	++
Tendenz zur Überanpassung	+	+	– –	–	++
Verständlichkeit für den Menschen	–	++	– –	++	– –
Legende: ++ sehr vorteilhaft, + vorteilhaft, – nachteilig, – – sehr nachteilig					

Tabelle 3-1: Zusammenfassung von Eigenschaften ausgewählter Methoden für Prädiktionsmodell

Die Tabelle 3-1 fasst die Eigenschaften der Verfahrensgruppen noch einmal zusammen. Der Einblick in die beschriebenen fünf Verfahrensgruppen zeigt, welche unterschiedlichen Herangehensweisen existieren sowie ihre Vor- und Nachteile. Anhand dieser Beurteilung werden im Verlauf der Arbeit Verfahren ausgewählt, die zur Problem- und Datenstruktur passen. Wie bereits erwähnt, existieren daneben noch zahlreiche andere Verfahren. Deren Anwendung ist allerdings häufig an bestimmte Voraussetzungen geknüpft, wie etwa normalverteilte Attributwerte, stationäre Signale oder lineare Zusammenhänge (vgl. [Lee10]). Beispiel für solche Verfahren sind Regression, Kalman Filter, Autoregressive-Moving-Average (ARMA), Hidden Markov Modell oder Match Matrix. Sie wurden hier nicht betrachtet, da ihre Voraussetzungen im vorliegenden Problemfall nicht erfüllt sind, wie in den Kapiteln 5 und 6 erläutert wird.

3.3 Metriken zur Qualitätsbeurteilung von Prädiktionsmodellen und Klassifikatoren

Nachdem im Abschnitt 3.1 der KDD-Prozess zur Orientierung für die Prädiktionsmodellentwicklung eingeführt und im Abschnitt 3.2 Prädiktionsverfahren verglichen wurden, wird nun die Qualitätsbeurteilung ausführlich diskutiert. Wenn es unmöglich ist, im Vorfeld einer Prädiktion zu bestimmen, welches Verfahren die besten Ergebnisse liefern wird (vgl. Abschnitt 3.2 und *No Free Lunch Theorem* von Duda et al. [DHS12]), braucht es Wege, um die Ergebnisqualität bewerten zu können. Im KDD spielen deshalb Performance Metriken, auch als Gütekriterien bekannt, eine wichtige Rolle. In Bezug auf die Prädiktion geht es im Wesentlichen darum, zu bewerten, wie gut die Hypothese aus der Vorhersage und die Realität übereinstimmen. Dazu werden zunächst Metriken bei binären Klassen (Abschnitt 3.3.1) und danach für multiple Klassen (Abschnitt 3.3.2) behandelt. Abschließend wird auf unterschiedliche Kosten einer Fehleinschätzung eingegangen (Abschnitt 3.3.3), da dies für die vorliegende Arbeit von besonderer Bedeutung ist.

3.3.1 Metriken bei der Klassifikation binärer Klassen

Im einfachsten Fall liegt ein binäres Klassifikationsproblem vor. Eine Klasse beinhaltet alle Objekte in funktionsfähigem Zustand, die andere Klasse dementsprechend alle Objekte mit Instandhaltungsbedarf. Für das Data Mining wird der relevante, aufbereitete Datensatz in eine oder mehrere Trainingsmengen und eine oder mehrere Testmengen geteilt. Ein Verfahren, z.B. eines der in Abschnitt 3.2 beschriebenen, wird verwendet, um anhand der Trainingsmenge eine Abbildungsfunktion der (Mess-)Daten auf die Klassen zu erlernen. Anschließend wird das trainierte Verfahren auf die Testmenge angewendet. Dadurch werden Daten klassifiziert, d.h. Hypothesen der jeweiligen Klassenzugehörigkeit aufgestellt, die mit der tatsächlichen Klassenzugehörigkeit verglichen werden können. Mit einer Konfusionsmatrix wird ein solcher Vergleich charakterisiert, wenn dieser die Anzahl an Objekten je hypothetischer und tatsächlicher Klasse gegenüberstellt. Wie in der Literatur fest definiert (z.B. in [Ega75, HM82, Swe88, Han97, Bra97, BBC⁺00]) und in Abbildung 3-7 dargestellt, ergeben sich bei binären Klassen die folgenden vier Fälle bzw. Felder in der Konfusionsmatrix:

- **richtig Positive (engl. True Positives) (TP)** – Anzahl der Objekte, die tatsächlich gestört sind und die als gestört klassifiziert wurden
- **falsch Negative (engl. False Negatives) (FN)** – Anzahl der Objekte, die tatsächlich gestört sind, aber als funktionsfähig klassifiziert wurden
- **falsch Positive (engl. False Positives) (FP)** – Anzahl der Objekte, die tatsächlich funktionieren, aber als gestört klassifiziert wurden
- **richtig Negative (engl. True Negatives) (TN)** – Anzahl der Objekte, die tatsächlich funktionieren und auch als funktionierend klassifiziert wurden

Nahezu alle der zahlreichen Performance Metriken lassen sich aus der Konfusionsmatrix berechnen. Sie reagieren jedoch unterschiedlich auf bestimmte Eigenschaften der zugrundeliegenden Datenbasis, bspw. auf die Datensatzgröße, die Klassenanzahl oder die Klassengröße. Es gibt daher Metriken, die grundsätzlich schlecht für die Bewertung von Prognosen bzw. Klassifikatoren geeignet sind. Aber es existieren auch Metriken, die abhängig vom Vorhersageziel, eine gute

		Hypothetischer Zustand	
		P := Anzahl Positive	N := Anzahl Negative
Tatsächlicher Zustand	T := Anzahl Fehler	TP := True Positives (richtig diagnostizierte Fehler)	FN := False Negatives (nicht diagnostizierte Fehler)
	F := Anzahl Nicht-Fehler	FP := False Positives (fälschlich diagnostizierte Fehler, Fehlalarme)	TN := True Negatives (richtig diagnostizierte Nicht-Fehler)

Abbildung 3-7: Struktur der Konfusionsmatrix bei binärer Klassifikation

oder schlechte Bewertungsbasis darstellen. Es soll auf einige Metriken und ihre Eigenschaften eingegangen werden, um im Verlauf der Arbeit geeignete Metriken zur Bewertung eines Prognosemodells für Weichen auswählen zu können. Zugunsten eines besseren Leseflusses wird hier weitestgehend darauf verzichtet, die Berechnungsformel aus den jeweiligen Quellen wiederzugeben. Diese befinden sich im Anhang A.2.

Die **relative Genauigkeit (engl. Accuracy) (Acc)** gehört zu den beliebtesten und meist verwendeten Performance Metriken, insbesondere bei Klassifikationsproblemen außerhalb der Informatik und Mathematik. Das liegt hauptsächlich an ihrer intuitiven Berechnung. Sie ergibt sich aus dem Verhältnis der Anzahl richtig klassifizierter Objekte zur Anzahl aller Objekte (vgl. Gleichung (A.11) im Anhang A.2). Die Acc ist leicht verständlich, und als skalarer Wert kann sie einfach für den Vergleich und die Rangfolge von Klassifikatoren genutzt werden. Allerdings besitzt die Acc auch einige gravierende Nachteile. Vor allem ihre Abhängigkeit von der Klassenverteilung ist der Grund dafür, dass sie im KDD-Prozess nur selten oder allenfalls zusätzlich zu weiterentwickelten Metriken verwendetet wird. [Han97, PFK98] Zwei einfache Beispiele verdeutlichen diese Schwäche. Die Konfusionsmatrizen 3-2 und 3-3 sind das Ergebnis desselben Klassifikators. Die Klassengröße unterscheidet sich jeweils um den Faktor 1000. Die Acc des ersten Beispiels unterscheidet sich mit 0.6 ($\frac{3000+12}{3000+2000+3+12} = 0.6$) deutlich von dem Wert 0.8 ($\frac{3+12000}{3+2+3000+12000} = 0.8$) des zweiten Beispiels, obwohl der Klassifikator ein und derselbe ist. Es wäre also falsch daraus zu schlussfolgern, dass der Klassifikator des einen Beispiels dem des anderen überlegen ist. Diese Varianz gegenüber der Klassenverteilung ist nachteilig, da die zukünftige Klassenverteilung im operativen Betrieb eines Prognosemodells unbekannt ist. Warum sollte daher eine Metrik zur Bewertung genutzt werden, die möglicherweise eine nicht vorhandene Überlegenheit eines Modells suggeriert? Deshalb sollte eine Performance Metrik unabhängig von der Klassenverteilung sein [Han97, PFK98, DH06].

	P	N
T	3000	2000
F	3	12

Tabelle 3-2: Beispiel 1 einer Konfusionsmatrix

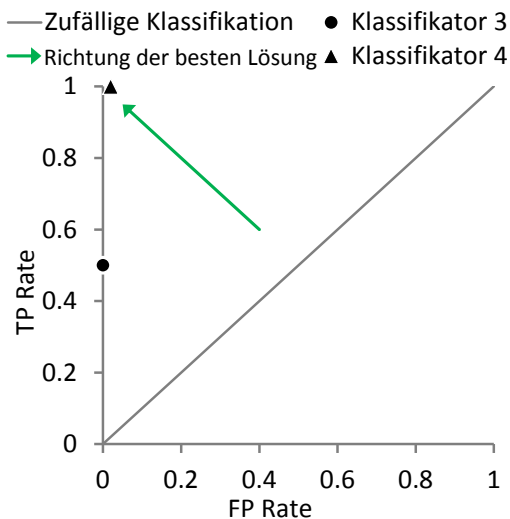
	P	N
T	3	2
F	3000	12000

Tabelle 3-3: Beispiel 2 einer Konfusionsmatrix

Andere populäre, aber nicht so häufig verwendete Metriken sind die Sensitivität (engl. Sensitivity oder Recall), die Relevanz (engl. Precision), die Spezifität (engl. Specifity) und der F-Wert (engl. F-score). Sie sind ebenfalls recht einfach aus der Konfusionsmatrix zu berechnen. Sokolova und Lapalme [SL09] haben systematisch untersucht, inwieweit diese Metriken invariant gegenüber Veränderungen in der Konfusionsmatrix sind. Ändert sich nur einer der vier Werte

aus der Konfusionsmatrix, weisen alle diese Metriken mindestens eine Invarianz auf. So zeigt z.B. die Relevanz-Metrik eine Veränderung der TN nicht an. Im Gegensatz zur Invarianz der Klassenverteilung, ist dies nicht per se gut oder schlecht. So ist bspw. bei der Begriffssuche im Internet die Zahl der irrelevanten Seiten so hoch, dass eine Bewertung der nicht gefundenen, irrelevanten Seiten (TN) außer Acht gelassen werden kann. Vielmehr orientiert sich die Wahl des besten Suchverfahrens am Verhältnis der gefundenen, relevanten Seiten (TP) zu allen gefundenen Seiten (TP und FP). Da alle diese Metriken bestimmte Teile der Konfusionsmatrix unberücksichtigt lassen, sollten sie im Allgemeinen nicht einzeln verwendet werden. Die Relevanz und der F-Wert sind zudem variant gegenüber der Klassenverteilung [SL09]. Sie sollten daher nicht zur Bewertung verwendet werden, analog zur Acc.

Der **Receiver Operating Characteristic (ROC)-Graph** ist eine Alternative zu den genannten Metriken. Er wurde 1975 von Egan [Ega75] eingeführt und 1988 von Swets [Swe88] auf die Bewertung und die Visualisierung der Performance von Klassifikatoren übertragen. Aus der Konfusionsmatrix wird zum einen die **FP Rate** berechnet, also der Anteil der falsch Positiven an allen funktionsfähigen Objekten (FP und TN, siehe Gleichung (A.16) im Anhang A.2). Die FP Rate wird darum auch als Fehlalarmrate bezeichnet. Zum anderen wird die **TP Rate** berechnet, also der Anteil der richtig Positiven an allen gestörten Objekten (TP und FN, siehe Gleichung (A.17) im Anhang A.2). Die TP Rate ist auch als Trefferrate bekannt. Der ROC-Graph ist eine deutlich bessere Metrik als die Acc, da er deren Nachteile überwindet, aber trotzdem sehr einfach zu bestimmen ist [Han97, PFK98, Faw03]. Darüber hinaus ist der ROC-Graph auch eine Visualisierungsmöglichkeit für die Performance. In dem Diagramm wird die FP Rate auf der Abszisse und die TP Rate auf der Ordinate abgetragen (siehe Abbildung 3-8).



Klassifikator 3

	P	N
T	1	1
F	0	100

Klassifikator 4

	P	N
T	2	0
F	2	98

Abbildung 3-8: ROC-Graph mit zwei beispielhaften Klassifikatoren

Tabelle 3-4: Konfusionsmatrizen der Klassifikatoren aus Abbildung 3-8

In der Abbildung 3-8 ist beispielhaft die Performance von zwei Klassifikatoren eingezeichnet. Klassifikator 4 befindet sich dicht an der linken, oberen Ecke und ist daher deutlich überlegen. Dort ist die Trefferrate hoch und die Fehlalarmrate gleichzeitig niedrig. Würde gemäß der jeweiligen Acc bewertet, entstünde ein gegenteiliger Eindruck, denn mit 0.99 ist die Acc des Klassifikators 3 höher als die 0.98 des Klassifikators 4. Ein Blick auf die dazugehörigen Konfusionsmatrizen in Tabelle 3-4 zeigt, dass die Klassen sehr unausgeglich sind. Den 100 negativen

Objekten stehen nur zwei positive gegenüber. Das offenbart eine weitere Schwäche der Acc, die ein ROC-Graph nicht besitzt [Faw03]. Wenn die Objekte sehr unausgeglichen über die Klassen verteilt sind, bevorzugt die Acc Klassifikatoren mit vielen Treffern in der Klasse mit sehr vielen Objekten [PFK98]. Aber gerade die seltenen Klassen sind meist die interessantesten, bspw. die wenigen Weichenstörungen, die es zu verhindern gilt.

Es sei ergänzt, dass in einem ROC-Graph Klassifikatoren meist nicht nur als Punkt abgebildet werden, sondern als Kurve. Klassifikatoren ermitteln für jedes Objekt einen Klassenzugehörigkeitswert. Je nachdem, ab welchem Niveau der Klassifikator ein Objekt der einen oder anderen Klasse zuordnet, ergeben sich dann unterschiedliche Konfusionsmatrizen. D.h. für jedes unterschiedliche Niveau entsteht eine Konfusionsmatrix, also auch unterschiedlich viele Punkte im ROC-Graph. Werden diese Punkte verbunden, ergeben sie die ROC-Kurve des Klassifikators (vgl. Abbildung A-1 in Anhang A.2).

Der Nachteil des ROC-Graphs ist, dass es schwerer ist, Klassifikatoren gemäß ihrer Performance in eine Rangfolge zu bringen. Dies ist bei zwei Werten (FP Rate und TP Rate) nicht so offensichtlich, wie bei einer skalaren Metrik. Aus diesem Grund wurde 1982 die **Fläche unter der ROC-Kurve (engl. Area Under the ROC Curve) (AUC)** von Hanley und McNeil [HM82] eingesetzt und diskutiert. Im wesentlichen beschreibt die AUC, wie gut ein Klassifikator in der Lage ist, zwei Klassen voneinander zu unterscheiden. Mathematisch sind die AUC, der Gini Index und die Wilcoxon Teststatistik gleich, wie Adams und Hand [AH99] gezeigt haben. Bradley [Bra97] hat mit der AUC nicht nur mehrere Klassifikationsverfahren bewertet, sondern auch ihre Eigenschaften und ihre Vorteile gegenüber der Acc beschrieben. Die AUC ist invariant gegenüber der Klassenverteilung und hängt nicht vom Niveau des Klassenzugehörigkeitswertes ab. Das ist vor allem wichtig, da die Klassenverteilung und das passende Niveau des Klassenzugehörigkeitswertes für den zukünftigen, operativen Betrieb unbekannt sind. Die AUC wird deshalb als skalare Performance Metrik für binäre Klassifikatoren empfohlen (vgl. [Bra97, AH99, HT01]).

3.3.2 Metriken bei der Klassifikation multipler Klassen

Der vorherige Abschnitt behandelte Metriken zur Qualitätsbewertung von binären Klassifikatoren. Dass solche Metriken auch bei der Evaluation der Zustandsprognose angewendet werden, haben z.B. Saxena et al. [SCB⁺08] und Leao et al. [LYRF08] gezeigt. Die Konzepte der Konfusionsmatrix und von Metriken lassen sich an binären Klassen zwar gut erläutern, aber für die Zustandsprognose in Form der RUL sind zwei Klassen zu grob. Doch die Anzahl der Klassen lässt sich theoretisch gegen unendlich steigern. Dadurch entsteht aus einer diskreten Einteilung in Klassen ein stetiger Wertebereich, der auch für die Vorhersage der verbleibenden Nutzungsdauer geeignet ist. Im folgenden Abschnitt wird deshalb die Thematik der Performance Metriken auf multiple Klassen erweitert.

Auch bei multiplen Klassen ist die Konfusionsmatrix der Ausgangspunkt (vgl. Abbildung 3-7). Allerdings sind die Felder nicht mehr eindeutig mit TP, etc. bezeichnet, sondern die Zeilenanfänge enthalten die tatsächlichen Klassen und die Spaltenanfänge die hypothetischen Klassen. Die Felder der Matrix enthalten aber nach wie vor die Anzahl der Objekte, welche der jeweiligen tatsächlichen Klasse angehören und der jeweiligen hypothetischen Klasse zugeordnet worden (siehe Tabelle 3-5).

Hypothetische Klasse Tatsächliche Klasse	C_{I1}	C_{I2}	\dots	C_{Ij}
C_{1J}	c_{11}	c_{12}	\dots	c_{1j}
C_{2J}	c_{21}	c_{22}	\dots	c_{2j}
\vdots	\vdots	\vdots	\ddots	\vdots
C_{IJ}	c_{i1}	c_{i2}	\dots	c_{ij}

Tabelle 3-5: Struktur der Konfusionsmatrix bei multipler Klassifikation

Um die Acc zu bestimmen, muss lediglich die Summe der Diagonalen durch die Summe aller Zellen geteilt werden. Denn entlang der Diagonalen befinden sich die richtigen Klassifizierungen, jedoch ohne dass diese im Vorfeld mit TP oder TN bezeichnet wurden. Für eine konsistente Sichtweise im Vergleich zur binären Klassifikation kann sukzessive jede diagonale Zelle einzeln als TP und die restlichen diagonalen Zellen als TN angesehen werden. Die übrigen Zellen der Zeile bzw. Spalte werden dementsprechend als FN bzw. FP angesehen. Dies ist in den Tabellen 3-6 und 3-7 veranschaulicht.

C_I	C_J	C_{I1}	C_{I2}	C_{I3}	C_{I4}
C_{1J}		TP_{11}	FN_{12}	FN_{13}	FN_{14}
C_{2J}		FP_{21}	TN_{22}	c_{23}	c_{24}
C_{3J}		FP_{31}	c_{32}	TN_{33}	c_{34}
C_{4J}		FP_{41}	c_{42}	c_{43}	TN_{44}

C_I	C_J	C_{I1}	C_{I2}	C_{I3}	C_{I4}
C_{1J}		TN_{11}	c_{12}	c_{13}	FP_{14}
C_{2J}		c_{21}	TN_{22}	c_{23}	FP_{24}
C_{3J}		c_{31}	c_{32}	TN_{33}	FP_{34}
C_{4J}		FN_{41}	FN_{42}	FN_{43}	TP_{44}

Tabelle 3-6: Konfusionsmatrix bei multipler Klassifikation mit der ersten Klasse als positive Klasse**Tabelle 3-7:** Konfusionsmatrix bei multipler Klassifikation mit der letzten Klasse als positive Klasse

Es ergeben sich nun zwei Wege, um aus der Konfusionsmatrix Metriken zu berechnen. Bei der makroskopischen Sicht werden die Metriken schrittweise berechnet und anschließend ihre Summe durch die Anzahl der Klassen geteilt. Bei diesem Ansatz geht jede Klasse gleichgewichtig in die Metrik ein, unabhängig von der Anzahl ihrer Objekte. Bei der mikroskopischen Sicht werden erst sukzessive die Summen der Zellen gebildet und anschließend die Metrik aus den Summen errechnet. Auf diese Weise gehen Klassen mit vielen Objekten tendenziell stärker in die Metrik ein als Klassen mit wenigen Objekten. Ob die Berechnung makroskopisch oder mikroskopisch erfolgen soll, ist eine Entscheidung, die sich an der zugrundeliegenden Problemstruktur und dem Ziel der Bewertung ausrichtet. Jedenfalls lassen sich die genannten Metriken (Sensitivität, Relevanz, Spezifität und F-Wert) so berechnen. Die Invarianzeigenschaften, auch die der Acc, bleiben aber bestehen, wie Sokolova und Lapalme [SL09] bewiesen haben. Demzufolge bleiben auch die damit verbundenen Nachteile bestehen.

Jurman und Furlanello [JF10] schlagen die **Konfusionsentropie (engl. Confusion Entropy) (CEN)** und den **Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient) (MCC)** als Metriken vor. Letzterer wurde zwar schon 1975 von Matthews [Mat75] eingeführt, fand aber bisher nur wenig Beachtung [JF10]. In ihren Vergleichen der beiden Metriken mit der Acc zeigen sie, dass die CEN und der MCC sehr viel besser bei unausgeglichene Klassengrößen geeignet sind. Sie werden nicht so stark von einer hohen Objektanzahl in einer Klasse bestimmt. Außerdem reagieren beide deutlich sensibler auf kleine Veränderungen in der Konfusionsmatrix.

Die CEN ist sogar noch etwas genauer darin als der MCC, aber für den binären Fall mathematisch nicht definiert, also nicht bei binären Klassen nutzbar. So empfehlen die Autoren dann auch den MCC als Maß für die Qualität bei multipler Klassifikation, zumindest solange auf einen ROC-Graphen verzichtet werden muss. [JF10, JRFZ12] Invariant gegenüber der Klassenverteilung sind CEN und MCC nicht.

Was den ROC-Graph und die AUC angehen, haben Hand und Till [HT01] sowie Ferri et al. [FHOS03] gezeigt, dass eine Erweiterung auf multiple Klassen möglich ist. Da aber die ROC-Analyse sukzessive und paarweise durchgeführt wird, wächst die Anzahl der Vergleiche exponentiell. Deshalb haben Landgrebe und Duin [LD07] auch einen Algorithmus zur effizienten Berechnung vorgestellt. Grundsätzlich können selbstverständlich die Makro- bzw. die Mikro-FP Rate und -TP Rate aus der Konfusionsmatrix berechnet werden und somit auch in üblicher Art zweidimensional visualisiert werden. Es ist allerdings zu beachten, dass die mikroskopisch berechneten Varianten der FP Rate, der TP Rate und der AUC nicht länger invariant gegenüber einer veränderten Klassenverteilung sind.

Darüber hinaus existieren zahlreiche Metriken, die auf bestimmte Aspekte bei multiplen Klassen besser eingehen sollen als andere Metriken. Diese Metriken werden zwar unterschiedlich berechnet, aber bei der Anwendung bewerten sie Klassifikatoren sehr ähnlich, so dass sie nicht wirklich einen Vorteil bieten. Das wurde von Ferri et al. [FHOM09] nachgewiesen. Deshalb wird hier nicht weiter auf sie eingegangen.

Die Tabelle 3-8 fasst die relevanten Eigenschaften der Metriken zusammen. Im weiteren Verlauf der Arbeit soll sich die Auswahl einer oder mehrerer Metriken für die Bewertung eines Prognosemodells an diesen Eigenschaften orientieren. In Abhängigkeit von der Datenbasis, die dem Zustandsprognosemodell für Eisenbahnweichen zugrunde liegt, soll so nachvollziehbar eine geeignete Qualitätsbeurteilung erfolgen. Denn darüber sind sich viele Autoren einig, die Wahl der richtigen Metrik entscheidet über die Wahl des richtigen Prädiktionsmodells (vgl. [BBC⁺00, Faw03, DH06, SCB⁺08, FHOM09, SL09]).

Eigenschaft		Anwendbar für		Invariant gegenüber Klassenverteilung	Variant gegenüber jeder einzelnen Zellwertänderung in der Konfusionsmatrix	Geegnet bei Rauschen in tatsächlichen Klassen	Geegnet wenn selbste Klassen wichtig sind	kleine Unterschiede bei vielen Klassen anzuzeigen
Metrik	binäre Klassen	multiple Klassen						
Sensitivität	+	+	+	+	-	+	-	-
Relevanz	+	+	+	-	-	+	-	-
Spezifität	+	+	+	+	-	+	-	-
F-Wert	+	+	+	-	-	+	-	-
Acc	+	+	+	-	+	+	-	-
Kappa Statistik	+	+	+	-	+	+	-	-
arithmetisches und geometrisches Mittel der Trefferrate je Klasse	+	+	+	+	+	+	0	-
AUC Varianten mit a priori Klassenverteilung	+	+	+	-	+	0	+	0
AUC Varianten mit gleichverteilten Klassen	+	+	+	+	+	0	+	0
gewertete AUC	+	+	+	+	+	0	-	0
Wahrscheinlichkeits-AUC	+	+	+	+	+	0	0	0
makroskopisch durchschnittliche Wahrscheinlichkeitsrate	+	+	+	+	+	0	0	0
mittlere Wahrscheinlichkeitsrate	+	+	+	-	+	0	0	0
mittlere absolute Fehler	+	+	+	-	+	0	0	0
mittlere quadratische Fehler	+	+	+	-	+	0	+	0
LogLoss	+	+	+	-	+	-	-	+
Kalibrierverlust	+	+	+	+	+	0	0	-
Kalibrierung nach Intervallen	+	+	+	-	+	-	-	-
ROC-Graph	+	+	+	+	+	0	+	0
Makro-AUC	+	+	+	+	+	0	+	0
Mikro-AUC	+	+	+	-	+	0	+	0
CEN	-	+	+	-	+	0	0	+
MCC	+	+	+	-	+	0	0	+
Legende: - nein/schlecht, + ja/gut, 0 mittelmäßig								

Tabelle 3-8: Zusammenfassung der Eigenschaften verschiedenen Performance Metriken

3.3.3 Beurteilung bei unterschiedlichen Kosten einer falschen Klassifikation

Bisher wurde die Qualitätsbeurteilung von Klassifikatoren und Prognosemodellen unter der Annahme betrachtet, dass die Kosten einer falschen Klassifikation gleich groß sind. Das ist für eine neutrale, allgemeine Beurteilung auch richtig. Im speziellen Anwendungsfall, in dem ein reales Problem zu lösen ist, unterscheiden sich meistens die Auswirkungen einer Missklassifikation [Tur95]. So hat z.B. in der Medizin die Klassifikation eines Patienten als gesund, während er tatsächlich todkrank ist (FN), sehr viel schwerere Folgen als der umgekehrte Fall (FP). Im Profisport hat die falsch Positive (FP) Klassifikation einer Dopingprobe meist drastischere Konsequenzen als ein falsch negatives Ergebnis (FN). Zumindest aus Sicht des Athleten, denn für ihn führt ersteres ungerechterweise zu einer Disqualifikation, einer Berufssperre oder gar dem Karriereende. Wie unterschiedliche Kosten einer falschen Klassifikation in die Qualitätsbeurteilung einfließen, erläutert der folgende Abschnitt. Er rundet damit die allgemeine Betrachtung von Metriken ab.

Die Problematik wurde bereits in der Vergangenheit adressiert, bspw. durch die kostensensitive Klassifikatorenevaluation von Turney [Tur95], Regret Graphen von Hilden und Glasziou [HG96] oder Loss Difference Plots von Adams und Hand [AH99]. Es sind auch Klassifikationsalgorithmen verfügbar, die die Missklassifikationskosten bereits in der Trainingsphase berücksichtigen. Ein Beispiel dafür ist der häufig verwendete Meta Cost Algorithmus von Domingos [Dom99]. Allerdings verlangen Regret Graphen und Meta Cost eine genaue Angabe der jeweiligen Missklassifikationskosten, was in der Realität oftmals problematisch ist. Denn wer kann in der Trainingsphase eines Algorithmus genau beziffern, wie z.B. die Auswirkungen einer falsch klassifizierten Krankheit sind. Adams und Hand [AH99] haben deshalb vorgeschlagen, eine bestimmte Verteilung der Kosten durch Experten schätzen zu lassen, die mit der grundlegenden Fragestellung vertraut sind, also z.B. durch Ärzte.

Sind die Kostenwerte für die Felder der Konfusionsmatrix bekannt, können diese in nahezu jede Metrik einbezogen werden. Am Beispiel des ROC-Graphen wird deutlich, welche Auswirkung unterschiedliche Missklassifikationskosten auf die Beurteilung haben können. Die beiden Klassifikatoren in Abbildung 3-9 sind rein technisch gesehen gleich gut. Sie liegen auf einer Linie, die parallel zur Diagonalen verläuft. Je nach Kosten einer falschen Klassifikation ist jedoch der eine Klassifikator besser als der andere. Die Iso-Performance-Linie, entlang derer Klassifikatoren gleich gut sind, steigt nicht länger um 1. Ihr Anstieg hängt nun vom Verhältnis der Kosten für FP zu den Kosten für FN ab. Im Fall der abgebildeten Missklassifikationskosten im oberen Teil der Tabelle 3-9 sind die Kosten für FP doppelt so hoch wie die Kosten für FN. Das bedeutet auch, dass eine erhöhte FP Rate mit einer zweimal so hohen TP Rate einhergehen muss, damit das Ergebnis als gleich gut angesehen werden kann. In diesem Fall ist der Klassifikator 7 besser, da seine Gesamtkosten von 270 ($50 \cdot 0 + 50 \cdot 3 + 20 \cdot 6 + 80 \cdot 0$) kleiner sind als die 345 ($75 \cdot 0 + 25 \cdot 3 + 45 \cdot 6 + 55 \cdot 0$) des Klassifikators 8. Klassifikator 8 liegt rechts der Iso-Performance-Linie des Klassifikators 7. Im Fall des zweiten Beispiels für Missklassifikationskosten ist es genau umgekehrt. Da die Kosten für FP nur ein Fünftel der Kosten für FN betragen, ist hier der Klassifikator 8 überlegen. Klassifikator 7 liegt unterhalb der Iso-Performance-Linie von Klassifikator 8.

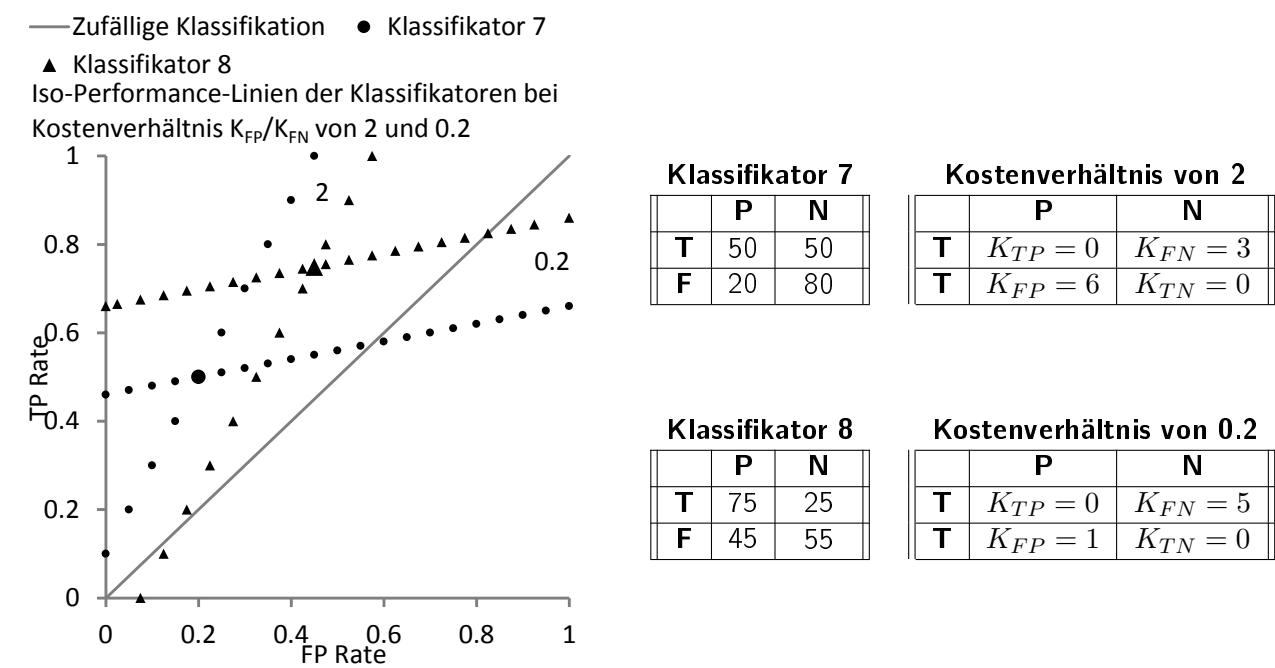


Abbildung 3-9: ROC-Graph mit zwei beispielhaften Klassifikatoren bei unterschiedlichen Missklassifikationskosten

Tabelle 3-9: Konfusionsmatrizen der Klassifikatoren und Kostenmatrizen der unterschiedlichen Missklassifikationskosten aus Abbildung 3-9

Die Betrachtung des Beispiels in Abbildung 3-9 gilt allerdings nur für die gegebenen Konfusions- und Kostenmatrizen in Tabelle 3-9. Ändert sich die tatsächliche Klassenverteilung, ändern sich auch die Iso-Performance-Linien. Diese Schwachstelle des ROC-Graphen bzw. von Metriken im Allgemeinen haben Drummond und Holte [DH06] ausführlich diskutiert. Sie zeigten, dass ROC-Graphen keine Aussage darüber erlauben, bei welchen Kosten und welcher Klassenverteilung ein Klassifikator besser ist als der andere. [DH06] Um diese kostensensitive Performance Evaluation vornehmen und visualisieren zu können, entwickelten die Autoren sogenannte Cost Curves. Bei Cost Curves werden die erwarteten Kosten eines Klassifikators als Funktion der Missklassifikationskosten und der Klassenverteilung abgebildet. Cost Curves decken dabei das gesamte Spektrum an möglichen Kosten und Klassenverteilungen ab. Demzufolge müssen bei Analysen mit Cost Curves weder exakte Kosten bekannt sein, noch eine Kostenverteilung geschätzt werden. Damit sind sie anderen Ansätzen überlegen. Ein weiterer Vorteil von Cost Curves ist ihr mathematischer Zusammenhang zu ROC-Graphen. Ein Punkt im ROC-Graphen ist eine Linie in der Cost Curves Visualisierung. Umgekehrt ist ein Punkt in der Cost Curve Visualisierung eine Iso-Performance-Linie im ROC-Graphen (vgl. [DH06]). Cost Curves und ROC-Graphen ergänzen einander. Sie bilden daher eine ausgezeichnete Kombination zur Qualitätsbeurteilung von Klassifikatoren, vor allem wenn die Missklassifikationskosten unbekannt sind.

Prinzipiell lassen sich alle genannten kostensensitiven Ansätze gleichermaßen auf binäre und multiple Klassifikationen anwenden. Missklassifikationskosten zu bestimmen und danach die Qualität zu beurteilen, wird allerdings aufwendiger, je mehr Klassen es zu klassifizieren gibt. Da für jedes Feld in der Konfusionsmatrix auch ein entsprechender Kostenwert gefunden werden muss, wächst die Anzahl der Kostenwerte im schlechtesten Fall quadratisch mit der Anzahl der Klassen [FHOS03, DH06]. Soll dann noch bei verschiedenen Kostenverteilungen der beste Klassifikator ermittelt werden, wird es sehr aufwendig, wie Landgrebe und Duin [LD07] argumentiert haben,

selbst mit ihrem effizienten Algorithmus. Außerdem nimmt die Unsicherheit zu. Je mehr Felder eine Konfusionsmatrix besitzt, für die eine Kostenverteilung geschätzt werden muss, desto höher ist die Wahrscheinlichkeit, insgesamt eine unrealistische Abschätzung zu verwenden.

Zusammenfassend lässt sich sagen, dass die Beurteilung bei unterschiedlichen Kosten einer falschen Prognose genauso von der zugrundeliegenden Problemstellung und der Datenstruktur abhängt, wie bei gleichen Missklassifikationskosten. Ob dabei die genauen Kosten, eine Kostenverteilungen oder Cost Curves eingesetzt werden, sollte der Anwender an der Verfügbarkeit zuverlässiger Information ausrichten. In den anschließenden Kapiteln, die die zu analysierende Datenbasis und mathematische Problemdefinition erläutern, werden daher die hier beschriebenen Grundlagen angewendet, um nachvollziehbar geeignete Metriken auszuwählen und die Bewertung der Prognose transparent darzulegen.

3.4 Zusammenfassung des Kapitels

In diesem Kapitel wurden Verfahren der Datenanalyse und der Zustandsprädiktion behandelt, um einen Überblick zu möglichen Herangehensweisen zu geben. Dabei wurde auf das Prognose und Zustandsmanagement (PHM) eingegangen. Dessen datengetriebene Ansätze sind erfolgversprechender als die zuverlässigkeits- oder modellbasierten Ansätze, weil erstere zu wenig die individuellen Eigenschaften abbilden, aber für letztere die Weiche und ihre Einflussfaktoren zu komplex für eine Modellierung sind. Datengetriebene Ansätze wurden mit der Wissensentdeckung in Datenbanken (KDD) verglichen. Im Ergebnis ist die Orientierung an der KDD für die Entwicklung eines Prognosemodells besser geeignet, da heterogene, umfangreiche Datenquellen zu berücksichtigen sind, versteckte Zusammenhänge identifiziert werden müssen und geeignete Evaluationsmethoden zur Ergebnisbewertung notwendig sind. Diese Punkte werden von der KDD besser abgedeckt als vom PHM.

Des Weiteren wurden einige, in der KDD bzw. im Data Mining häufig verwendete Verfahrensguppen beschrieben, mit denen sich die Ziele dieser Arbeit erreichen lassen. Für k-nächste-Nachbarn (kNN), Bayes'sche Netze, künstliche Neuronale Netze (aNN), Entscheidungsbaumverfahren (DTree) und Support Vector Machines (SVM) wurden die grundlegenden Funktionen sowie die Vor- und Nachteile erläutert. Auf diese Beurteilung wird im Verlauf dieser Arbeit zurückgegriffen, wenn ein oder mehrere geeignete Verfahren auszuwählen sind.

Zusätzlich wurden Metriken zur Qualitätsbeurteilung von Prädiktionsmodellen und Klassifikatoren diskutiert. Sie ermöglichen es, die Performance der verschiedenen Prognoseverfahren miteinander zu vergleichen. Doch die Metriken reagieren unterschiedlich auf bestimmte Eigenschaften der zugrundeliegenden Daten- und Problemstruktur, etwa die Klassenanzahl, die Klassenverteilung oder die Klassengröße. Daher sollte für die Evaluation eines Prognosemodells eine geeignete Metrik gewählt werden. Denn die Wahl der richtigen Metrik ist die Voraussetzung für die Wahl des richtigen Prognosemodells.

4 Anforderungen an das zu entwickelnde Vorhersagemodell des Weichenzustands

Das Kapitel 2 behandelte die Weiche und ihre Funktion, ging auf Instandhaltung im Allgemeinen und auf zustandsorientierte Instandhaltung im Speziellen ein. Daraus wurde insbesondere deutlich, dass eine verlässliche Prognose des Zustands fehlt, um Weichen effizienter instandzuhalten sowie zusätzlich die hohe Zahl an Verspätungsminuten zu reduzieren. Bestehende Prädiktionsmodelle für Weichen bzw. Weichenkomponenten und ihre Anwendbarkeit wurden diskutiert, mit der Schlussfolgerung, dass keines der Modelle eine rechtzeitige oder verlässliche Prognose des Abnutzungsvorrats oder der verbleibenden Nutzungsdauer (engl. Remaining Useful Life) (RUL) für Weichen erlaubt. Kapitel 3 griff dies auf, indem Verfahren der Datenanalyse und der Zustandsprediktion verglichen wurden, die Teil eines Zustandsprognosemodells für Eisenbahnweichen sein können. Es wurde argumentiert, wie solche Verfahren hinsichtlich Qualität und Performance evaluiert werden können. Einige der Kritikpunkte zu den bestehenden Prädiktionsmodellen werden in diesem Kapitel aufgegriffen, um daraus Anforderungen an das zu entwickelnde Prognosemodell abzuleiten. Diese spiegeln auch Anforderungen wider, die sich aus den Zielen zur Beantwortung der Forschungsfragen der Arbeit ergeben (siehe Abschnitt 1.2). Zunächst werden allgemeine Anforderungen (Abschnitt 4.1) beschrieben und anschließend Anforderungen an die Genauigkeit (Abschnitt 4.2). Insgesamt soll die Erfüllung der Anforderungen sicherstellen, dass das in dieser Arbeit zu entwickelnde Prognosemodell effektiv, praktikabel und wirtschaftlich ist.

4.1 Allgemeine Anforderungen

Zu den allgemeinen Anforderungen zählen vor allem qualitative Anforderungen, die sich nur schwer in einem messbaren Wert ausdrücken lassen. Darunter fällt eine Anforderung, die in dieser Arbeit schon mehrfach genannt wurde (siehe Abschnitt 2.3.1 und 3.1.2) und auch in der Literatur häufig zu finden ist (vgl. [ZAB⁺02, Hes02, GR06, GB11, BG11, ARK13]): Sensoren, die im Gleisbett eingesetzt werden sollen, müssten sehr robust gegenüber Schmutz, Erschütterungen, etc. sein. Das macht sie meist auch teuer. Gleichzeitig ist jeder Sensor per se eine Fehlerquelle, die ausfallen kann und Wartung benötigt. Darum wird aus wirtschaftlichen und technischen Gründen die Anforderung gestellt, so wenig Sensoren wie möglich einzusetzen.

Mitunter wurden bei den bestehenden Prognosemodellen Sensoren verwendet, die nicht für die Anwendung im regulären Bahnbetrieb zugelassen sind (siehe Abschnitt 2.4). Ohne den enormen Aufwand einer Zulassung durch das Eisenbahn-Bundesamt (EBA), können solche Sensoren nicht genutzt werden. Auch wenn es theoretisch möglich wäre, bliebe das Ziel, Verspätungen zu verhindern praktisch unerreicht. Oder es wäre mit extrem hohen Kosten verbunden, die die Wirtschaftlichkeit infrage stellen würden. Darum sollte das zu entwickelnde Prognosemodell auf Daten von Sensoren basieren, die für die Anwendung im regulären Bahnbetrieb zugelassen sind, bspw. weil sie nachweislich rückwirkungsfrei funktionieren.

Wie in Abschnitt 2.1.2 beschrieben, befinden sich im deutschen Eisenbahnnetz Weichen mit sehr unterschiedlichen Eigenschaften. Damit das Prognosemodell eine gewisse Allgemeingültig-

keit besitzt, sollen Weichen mit unterschiedlichen Eigenschaften in die Entwicklung eingebunden werden, z.B. verschiedene Weichentypen, -verschlüsse, -radien. Denn mangelnde Allgemeingültigkeit ist schon bei den existierenden Diagnosemodellen ein Problem für deren Anwendbarkeit. Erfüllt das zu entwickelnde Modell diese Anforderung, hebt es sich zusätzlich von bestehenden Ansätzen ab, selbst wenn diese noch gar keine Prognose ermöglichen.

Bei einigen der bestehenden Prognosemodelle wurde kritisiert, dass nur ein einzelnes oder nur sehr wenige Objekte verwendet wurden. Die Allgemeingültigkeit ist dadurch nicht gegeben oder lässt sich anzweifeln. Daher soll eine größere Anzahl von Weichen in die Entwicklung des Prognosemodells eingebunden werden. Eine genaue Zahl lässt sich nicht begründet nennen, aber mehr als zehn erscheint notwendig, um zufällige Einzelbeobachtungen ausgrenzen zu können.

Des Weiteren wird gefordert, bei der Entwicklung des Prognosemodells Weichen und Antriebe zu verwenden, die im deutschen Eisenbahnnetz häufig verbaut sind. Denn die vorliegende Arbeit zielt ja gerade darauf ab, die Pünktlichkeit und die Wirtschaftlichkeit in Deutschland zu erhöhen.

Ein anderer Kritikpunkt an vielen bestehenden, vermeintlichen Prognosemodellen ist die zugrundeliegende künstliche Datenbasis, die experimentell erhoben wurde (siehe Tabelle 2-2). Damit ein Prognosemodell möglichst einfach übertragen und im realen Betrieb angewendet werden kann, sollte seine Entwicklung auch auf realen Messdaten basieren. Dies gilt sowohl für die Zustandsdaten als auch für die Störungsdaten.

Wenn die Datenbasis möglichst die Realität abbilden soll, muss diese auch ein gewisses Spektrum an Störereignissen und vor allem verschiedenen Umweltbedingungen abdecken, z.B. heiße, kalte, sonnige, regnerische und verschneite Tage. Daher sollte die Datenbasis mindestens einen Zeitraum von einem Jahr umfassen.

Wie wichtig die Erfüllung dieser Anforderungen für eine realistische Zustandsprognose ist, zeigen auch die Verbesserungsvorschläge, welche die Entwickler bestehender Prognosemodelle bei ihren eigenen Ansätzen sehen. So schlug z.B. Asada am Ende seiner Dissertation vor, die Ansätze mit realen Felddaten aus einem längeren Zeitraum und von verschiedenen Weichenantrieben zu wiederholen und zu überprüfen, bevor ein operativer Einsatz ausprobiert wird [Asa13, S. 143 ff.].

4.2 Genauigkeitsanforderungen an das Prognosemodell aus wirtschaftlicher Sicht

Zu den allgemeinen Anforderungen des vorherigen Abschnitts kommt eine essentielle Forderung hinzu: die Prädiktion des Zustands in Form einer verbleibenden Nutzungsdauer (RUL). Denn nur so kann die Instandhaltung Maßnahmen planen und Ressourcen bereitstellen. Dabei ist die Zustandsprädiktion im Grunde ein Mittel, um die Wirtschaftlichkeit einer Anlage zu erhöhen, indem die Kosten der Instandhaltung gesenkt werden (siehe Abschnitt 2.2.4). Dies ist allerdings kein gegebener Grundsatz. Vielmehr hängen die Senkung von Instandhaltungskosten und somit die Wirtschaftlichkeit eines Prognosemodells von der Genauigkeit seiner Vorhersage ab. Ein System, das zu viele Fehlalarme erzeugt oder zu wenig Störungen rechtzeitig erkennt, ist unwirtschaftlich. Denn es induziert unter Umständen mehr Kosten durch unnötige Instandhaltungsmaßnahmen, als es Kosten durch verhinderte Ausfälle einspart.

Natürlich ist aus technischer Sicht immer ein perfektes System anzustreben, d.h. ohne Fehlalarme und mit 100 Prozent richtig vorhergesagten Störungen. Als Metrik ausgedrückt, ist dies gleichbedeutend mit dem Punkt am linken, oberen Rand des Receiver Operating Characteristic (ROC)-Graphs, also einer falsch Positiv (FP) Rate von 0 und einer richtig Positiv (TP) Rate von 1 (siehe Abschnitt 3.3.1 bzw. Abbildung 3-8). Es ist fast unmöglich, diese Perfektion in der Realität zu erreichen. Das ist aber auch nicht notwendig. Es genügt, wenn die Kosten einer falschen Vorhersage geringer sind als der Nutzen durch eine richtige Vorhersage. Um dies bewerten und nachweisen zu können, wurde eine eigene Methode entwickelt, welche in den folgenden Abschnitten erläutert wird.

4.2.1 Transformation der stetigen Restnutzungsdauer auf eine binäre Klassifikation

Wie die Genauigkeit bzw. die Qualität einer Vorhersage beurteilt werden kann, wurde im Abschnitt 3.3 allgemein beschrieben. Insbesondere der Passus 3.3.3 erläutert, wie dabei der Nutzen einer richtigen bzw. die Kosten einer falschen Vorhersagen berücksichtigt werden können. Dazu bedarf es der Konfusionsmatrix und einer Matrix für die Missklassifikationskosten. Gegeben ist eine Menge von Messungen S und davon eine Menge an Samples \mathcal{S} einzelner Messungen s sowie eine Menge von Klassen Z , die den Zustand ausdrücken.

$$\mathcal{S} = \{s_l: 1 \leq l \leq S\} \quad (4.1)$$

Die Klassen bestehen aus kontinuierlichen Werten b der stetigen RUL und sind daher theoretisch unendlich bzw. Teil der reellen Zahlen \mathbb{B} im Intervall zwischen 0 und der maximalen Lebensdauer $maxRUL$.

$$Z: RUL \rightarrow \mathbb{B} \quad \text{mit } \mathbb{B} = \{b: 0 \leq b \leq maxRUL\}_{\mathbb{R}} \quad (4.2)$$

Es werden zwei Funktionen definiert, die jeweils den tatsächlichen Zustand TZ und den hypothetischen Zustand (aus der Prognose) HZ eines Samples beschreiben.

$$TZ: S \rightarrow Z \quad (4.3)$$

$$HZ: S \rightarrow Z \quad (4.4)$$

Die dazugehörige Konfusionsmatrix \mathbf{C} ist eine quadratische Matrix, ebenfalls theoretisch von unendlicher Größe.

$$\mathbf{C} \in \mathcal{M}(Z \times Z, \mathbb{B}) \quad \text{mit } C_{i,j} = |\{s \in \mathcal{S}: TZ(s) = i, HZ(s) = j\}| \quad (4.5)$$

Theoretisch müsste nun für jedes Feld der Konfusionsmatrix \mathbf{C} auch ein Kostenwert der Missklassifikationskosten oder des Prognosenutzens bestimmt werden. Die Kostenmatrix \mathbf{K} wäre damit theoretisch genauso unendlich groß.

$$\mathbf{K} \in \mathcal{M}(Z \times Z, \mathbb{B}) \quad \text{mit } K_{i,j} = K(C_{i,j}) \quad (4.6)$$

Die Prognose des Weichenzustands als stetige Funktion zur Abbildung der RUL zu betrachten, ist aber aus mehreren Gründen unpraktisch und unnötig:

1. Wie schon im Abschnitt 3.3.3 beschrieben, sind die Missklassifikationskosten meist nicht bekannt, erst recht nicht in einem stetigen Bereich. Angenommen, die tatsächliche RUL beträgt 100 Tage. Um welchen Kostenbetrag unterscheidet sich dann eine prognostizierte RUL von 30 Tagen von einer mit 40 Tagen oder 85.12 Tage von 85.13 Tagen?
2. Es wird eine Granularität der Prognose vermittelt, die bei Eisenbahnweichen mit großer Unsicherheit verbunden ist. Denn die Zustandsmessungen werden von vielen Faktoren beeinflusst und unterliegen starken Schwankungen. Das birgt wiederum die Gefahr, dass auch die RUL schwankt und das System unzuverlässig erscheint.
3. Für einen stetigen Wertebereich der Zustandsfunktion muss auch der Definitionsbereich, d.h. die Messfrequenz der zustandsbeschreibenden Parameter, stetig sein. In der Praxis ist der Unterschied zwischen einem Überwachungsintervall von z.B. 0.5 Sekunden (entspricht der maximalen Offenbarungszeit des signaltechnischen Zustands von Feldelementen gegenüber dem Stellwerk) und einer Minute vernachlässigbar gering. Das bedeutet, dass ein stetiger Definitionsbereich nicht effizient ist.
4. Um die Wirtschaftlichkeit der Zustandsprädiktion beurteilen zu können, ist es nicht von Bedeutung, ob eine RUL von einer oder 0.1 Stunde prognostiziert wird. Wichtig ist im Wesentlichen, ob die Störungsvorhersage rechtzeitig erfolgt, damit eine Störung verhindert werden kann.

Es stellt sich dann natürlich die Frage, was „*rechtzeitig*“ für die RUL-Prognose bedeutet. Das ist im Übrigen eine Frage, auf die auch die bisherigen Arbeiten zum Thema Weichenzustandsdiagnose oder -prognose keine Antwort geben konnten (siehe Abschnitt 2.4). Um sie zu beantworten, wurde deshalb im Rahmen der Forschung zur vorliegenden Arbeit ein Workshop mit Instandhaltungspersonal der DB Netz AG durchgeführt [Böh13c]. Dabei wurden Weichenmechaniker, LST-Meister und Instandhaltungsmanager befragt, wie lange eine Störung vorher erkannt werden müsste, um rechtzeitig Gegenmaßnahmen durchführen zu können – mit folgendem Ergebnis:

- Die **Notfallentstörung** muss **1 bis 2 Stunden** vor Störereignis ausgelöst werden. Diese Zeit wird in der Regel von den Instandhaltern benötigt, damit sich diese zur Weiche begeben und die Maßnahme durchführen können, eventuell auch während des Bereitschaftsdienstes. So ließe sich gerade noch unterbinden, dass der Zugverkehr beeinträchtigt wird, vorausgesetzt die Strecke muss nicht extra gesperrt werden.
- Für eine **geplante Maßnahme** muss die potenzielle Störung **3 bis 5 Tage** vorher bekannt sein. In der Regel ist dieser Zeitraum ausreichend, damit die Instandhaltungsmaßnahme in die Feinplanung für die Arbeitseinsätze entsprechend einfließen kann. Dadurch kann nahezu garantiert werden, dass der reguläre Zugbetrieb nicht beeinträchtigt wird.

Soll also der Bahnbetrieb ohne Weichenstörungen ablaufen, muss das zu entwickelnde Prognosemodell mindestens zwei Klassen identifizieren (siehe Abbildung 4-1). Die erste Klasse enthält Objekte, deren RUL fünf Tage oder weniger beträgt. Die zweite Klasse enthält alle Objekte, deren RUL größer ist. Es kann angenommen werden, dass die Kosten einer (falschen) Klassifikation sich nicht signifikant unterscheiden, wenn die RUL innerhalb der beiden Klassen weiter detailliert wird. Aus wirtschaftlicher Sicht reduziert sich die Prognose somit auf eine binäre Klassifikation. Damit ist zwar nicht gesagt, dass für das Instandhaltungspersonal zwei Klassen ebenfalls ausreichen. Denn für die grobe Planung, Aussagen zum Netzzustand oder aus Akzeptanzgründen

sind mehr Klassen interessant, aber eben nicht für eine Betrachtung der Wirtschaftlichkeit.

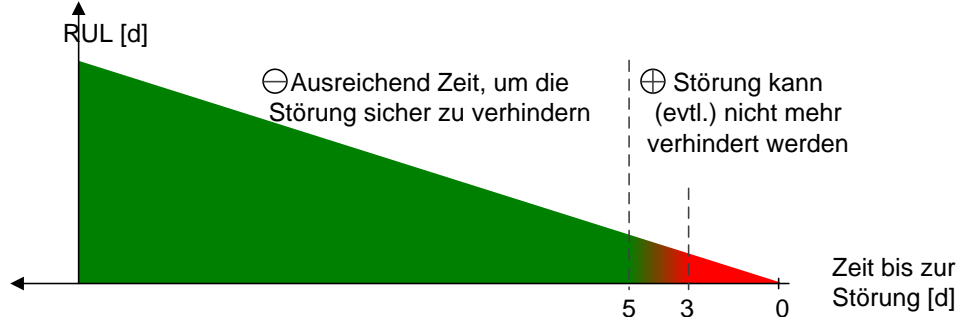


Abbildung 4-1: Reduktion der RUL-Prognose auf zwei Klassen auf Basis der Aussagen von Instandhaltern der DB Netz AG

Die Gleichung 4.2 wird daher wie folgt geändert:

$$Z: RUL \rightarrow \{\ominus, \oplus\} \quad \text{mit } z(RUL) = \begin{cases} \ominus: 5 < RUL \\ \oplus: 0 \leq RUL \leq 5 \end{cases} \quad (4.7)$$

Die Konfusionsmatrix erhält dadurch die Form $\mathbf{C} \in \mathcal{M}(2 \times 2)$. Außerdem können die bekannten Begriffe richtig Positiv, falsch Negativ, falsch Positiv und richtig Negativ verwendet werden (siehe auch Abbildung 3-7).

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{bmatrix} = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (4.8)$$

Für die binäre Konfusionsmatrix lässt sich nun verhältnismäßig einfach eine Kostenmatrix aufstellen, da diese lediglich vier klar definierte Fälle enthält.

$$\mathbf{K} = \begin{bmatrix} K_{TP} & K_{FN} \\ K_{FP} & K_{TN} \end{bmatrix} \quad (4.9)$$

Dabei sind die einzelnen Positionen, aus denen sich diese vier Kostenfälle zusammensetzen, vorerst weniger bedeutend für die Evaluation als die Kostenmatrix selbst. Daher wird zunächst die allgemeine Methode zur Genauigkeitsbewertung aus wirtschaftlicher Sicht hergeleitet und dann die Zusammenstellung der Kostenpositionen detailliert.

4.2.2 Break Even Situation einer Vorhersage

Während im vorangegangenen Abschnitt geklärt wurde, was „rechtzeitig“ bei der RUL-Prognose heißt und wie sie sich dadurch auf ein binäres Klassifikationsproblem reduziert, widmet sich der folgende Teil der wirtschaftlichen Performance. Ein Vorhersagemodell befindet sich im ökonomischen Gleichgewicht, wenn es durch richtig vorhergesagte Störungen (TP) genau so viel Kosten einspart, wie es durch Fehlalarme (FP) unnötiger Weise verursacht. Dieses Gleichgewicht wird hier als Break Even Situation bezeichnet (in Anlehnung an den Break Even Point bei der Produktion). Damit das Modell wirtschaftlich arbeitet, muss daher die Genauigkeit so hoch sein, dass der Nutzen größer ist als die Kosten. Es ist zu beachten, dass die Wirtschaftlichkeit der Vorhersagegenauigkeit nicht mit der Amortisation der Anschaffung eines Diagnose- und Prognosesystems verwechselt wird. Bei der Amortisation spielen vordergründig Anschaffungs- und Besitzkosten eine Rolle. García Márques et al. [GMLTR08] haben 2008 eine solche Amortisa-

tion für die zustandsorientierte Instandhaltung von Weichen berechnet. In ihren Betrachtungen nehmen sie allerdings nur theoretisch an, dass ein Diagnosesystem auch Störungen vermeiden könnte. Wie in Abschnitt 2.4 gezeigt wurde, existiert aber bisher keine Störungsvorhersage für Weichen. Obendrein verwenden sie nur einen konstanten Wert für die Fehlalarmrate (FP Rate) und gehen nicht weiter auf die Genauigkeit ein. Amortisationsrechnung und Vorhersagegenauigkeit sind zwei verschiedene Aspekte der Wirtschaftlichkeit, können aber kombiniert werden, wenn dazu Bedarf besteht.

Die Genauigkeit wird in Form einer Performance Metrik abgebildet. Als Metrik werden hier ROC-Graph bzw. die FP Rate (FP_r) und die TP Rate (TP_r) verwendet, da diese genau den Anteil an Fehlalarmen und an richtig prognostizierten Störungen ausdrücken. Außerdem sind sie invariant gegenüber der Klassenverteilung und eignen sich sehr gut für eine Visualisierung (vgl. Abschnitt 3.3.1 oder Tabelle 3-8). Um die Break Even Situation zu bestimmen, werden Metrik und Nutzenwerte gemeinsam verwendet.

$$FP_r = \frac{FP}{F} = \frac{FP}{FP + TN} \quad (4.10)$$

$$TP_r = \frac{TP}{T} = \frac{TP}{TP + FN} \quad (4.11)$$

Jede verhinderte Störung spart genau die Kosten ein, die ansonsten durch die Störung entstanden wäre. Mathematisch ausgedrückt entspricht der Nutzen von TP (B_{TP}) einfach den Kosten von FN mit positivem Vorzeichen. Die Kosten von FP werden ebenfalls als Nutzen ausgedrückt (B_{FP}). Dieser ist allerdings negativ. Werden nun die Kosten- und Nutzensätze mit den dazugehörigen Werten der Metrik und der Konfusionsmatrix multipliziert, ergibt sich die Wirtschaftlichkeit – der erwartete Nutzen B .

$$B = TP_r \cdot T \cdot B_{TP} + FP_r \cdot F \cdot B_{FP} \quad \text{mit } B_{TP} = K_{FN} \stackrel{!}{\geq} 0 \text{ und } B_{FP} = -K_{FP} \stackrel{!}{\leq} 0 \quad (4.12)$$

Das Ergebnis gilt allerdings nur für eine bestimmte Anzahl an Störungen (T) und Nicht-Störungen (F) sowie für einen bestimmten Nutzen- und Kostensatz. Auf diese Weise ließe sich nur ein ganz bestimmtes Szenario bewerten. Die Wirtschaftlichkeit soll jedoch für unterschiedliche Nutzen- und Kostensätze, sowie eine unterschiedliche Anzahl an Störungen nachweisbar sein. Letzteres ist insbesondere wichtig, da die zukünftige Störungshäufigkeit während der Trainings- und Testphase eines Klassifikators unbekannt ist. Außerdem variiert die Anzahl der Störungen und Messungen von Weiche zu Weiche. Die Prognose soll nicht nur für ein bestimmtes Szenario gut funktionieren, sondern allgemeingültige, gute Vorhersagen liefern. Um diesen Umstand zu berücksichtigen, könnten eigentlich Cost Curves, wie sie in Abschnitt 3.3.3 beschrieben wurden, angewendet werden. Bei Cost Curves werden die erwarteten Kosten eines Klassifikators als Funktion der Missklassifikationskosten (K_{FN}, K_{FP}) und der Klassenverteilung abgebildet. Obwohl Cost Curves eines der besten Bewertungsverfahren von Klassifikatoren bei unbekannten oder variierenden Missklassifikationskosten sind, eignen sie sich aus folgenden Gründen nicht für die vorliegende Fragestellung zur Break Even Situation. Zum einen sollen mit Cost Curves die Kosten für FN und FP minimiert werden, wie es für die allgemeine Klassifikatorenevaluation üblich ist. Für die Break Even Situation werden allerdings Nutzen von TP und Kosten von FP zusammen verwendet. Es ändert sich aufgrund der verschiedenen Vorzeichen die Art, wie Cost Curves normalisiert und interpretiert werden. Der beste Scorer ist nicht länger derjenige, welcher

Kosten von Null aufweist, sondern das Nutzenmaximum. Die originäre Interpretation ist nicht mehr gültig, da diese dazu führen kann, dass sogar ein Klassifikator gegenüber anderen favorisiert wird, der als einziger überhaupt keinen positiven Nutzen erzeugt. Zum anderen wird die eigentliche Performance der Klassifikatoren nicht dargestellt, weshalb zusätzlich noch der ROC-Graph gezeichnet werden muss. Und Cost Curves können die Break Even Situation nicht visualisieren. All diese Schwächen wurden bereits von Böhm [Böh13b] nachgewiesen. Im selben Artikel wurde auch eine eigene Herangehensweise entwickelt, um die Genauigkeit aus wirtschaftlicher Sicht zu beurteilen. Sie bildet den Kern zum Nachweis einer kostensparenden Prognose und wird deshalb im Folgenden ausführlich erläutert.

Im ersten Schritt wird der erwartete Nutzen aus Gleichung (4.12) bezüglich des relativen Anteils von Störungen (T_r) und Nicht-Störungen (F_r) an allen Messungen normalisiert. Es entsteht die erwartete Nutzenrate B_r .

$$\begin{aligned} B_r &= TP_r \cdot T_r \cdot B_{TP} + FP_r \cdot F_r \cdot B_{FP} & \text{mit } T_r = \frac{T}{T+F} \text{ und } F_r = \frac{F}{T+F} \\ &= TP_r \cdot T_r \cdot B_{TP} + FP_r \cdot (1 - T_r) \cdot B_{FP} & \text{und } T_r + F_r = 1 \Leftrightarrow F_r = 1 - T_r \\ &= (TP_r \cdot B_{TP} - FP_r \cdot B_{FP}) \cdot T_r + FP_r \cdot B_{FP} \end{aligned} \quad (4.13)$$

Hiermit ist nicht länger die genaue Anzahl an T und F relevant, sondern nur noch der Anteil an Störungen T_r . Außerdem werden alle möglichen Kombinationen von T und F auf das Intervall zwischen 0 und 1 abgebildet. Klassifikatoren werden somit vergleichbar, selbst wenn sie unterschiedliche Datensätze benutzen. Nun wird ein Punkt der FP Rate und der TP Rate P_0 definiert, der die Break Even Situation erreicht und gleichzeitig eine FP Rate größer 0 aufweist.

$$P_0(FP_r, TP_r) = \{(FP_r, TP_r) \in [0, 1]: FP_r > 0, B_r > 0\} \quad (4.14)$$

Durch Kombinieren und Umstellen der Gleichungen (4.13) und (4.14) wird es möglich, die geforderte Trefferrate in Abhängigkeit von der Fehlalarmrate und dem Störungsanteil darzustellen.

$$f(FP_r, T_r) = \frac{FP_r \cdot T_r \cdot B_{FP} - FP_r \cdot B_{FP}}{T_r \cdot B_{TP}} = TP_r \quad (4.15)$$

Dadurch wird eine Ebene im dreidimensionalen Raum definiert, die diesen zerteilt. Der Raum stellt den regulären ROC-Space aus FP Rate und TP Rate dar, erweitert um die Dimension des Störungsanteils. Für ein gegebenes Verhältnis von Fehlalarmkosten und Nutzen separiert die Ebene eine wirtschaftliche Performance von einer unwirtschaftlichen. Abbildung 4-2 zeigt die Ebene der Break Even Situation, in diesem Fall bei einem Verhältnis von $B_{FP}/B_{TP} = -0.5$.

Weil jedoch auch die Werte für die Kosten und den Nutzen unbekannt sind oder zumindest variieren können, muss diese Unsicherheit ebenfalls in der Bewertung berücksichtigt werden. Dazu wird das Kosten-Nutzen-Verhältnis (engl. Cost-Benefit-Ratio) (CBR) eingeführt und in die Gleichung (4.15) integriert.

$$f(FP_r, T_r, CBR) = \frac{(T_r - 1) \cdot FP_r \cdot CBR}{T_r} = TP_r \quad \text{mit } CBR = \frac{B_{FP}}{B_{TP}} \quad (4.16)$$

Analog zu Gleichung (4.15) entsteht so eine Hyperebene im vierdimensionalen Hyperraum, die

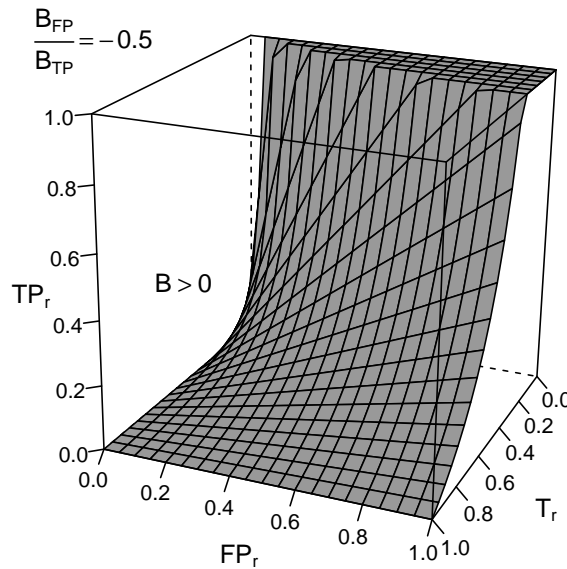


Abbildung 4-2: Beispiel für die Visualisierung der Break Even Situation bei gegebenem Verhältnis von Kosten eines Fehlalarms zum Nutzen einer verhinderten Störung

die Break Even Situation beschreibt. Und auch diese Ebene teilt den Raum in die Bereiche einer wirtschaftlichen und einer unwirtschaftlichen Genauigkeit. Die Visualisierung der vier Dimensionen ist zugegebener Maßen sehr schwierig, weshalb hier darauf verzichtet wird. Wenn jedoch ein minimales und ein maximales CBR bestimmt werden können, ist die Darstellung im dreidimensionalen Raum möglich. Angenommen, es ist bekannt, dass die Kosten eines Fehlalarms bzw. der Nutzen einer verhinderten Störung einen bestimmten Betrag nicht unterschreiten, aber auch einen bestimmten Betrag nicht übersteigen. Mit diesen Werten können dann zwei Ebenen definiert werden. Die eine basiert auf dem bestmöglichen CBR ($maxCBR$) und stellt die Mindestanforderung dar, damit ein Vorhersagemodell überhaupt wirtschaftlich sein kann. Die andere basiert auf dem schlechtesten CBR ($minCBR$) und bezeichnet die Grenze, ab der ein Vorhersagemodell in jedem Fall einen wirtschaftlichen Nutzen generiert. Das Prinzip ist in Abbildung 4-3 veranschaulicht. Würden Klassifikatoren darin abgebildet, müssten sie mindesten im sichtbaren Bereich zwischen beiden Ebenen sein, um vom ökonomischen Standpunkt aus eine ausreichende Genauigkeit aufzuweisen. In allen Fällen, in denen sie vor der transparenten Ebene liegen, erzeugen sie immer einen wirtschaftlichen Nutzen.

$$maxCBR = \frac{max(B_{FP})}{max(B_{TP})} \quad (4.17)$$

$$minCBR = \frac{min(B_{FP})}{min(B_{TP})} \quad (4.18)$$

Für die Evaluation der Genauigkeit kann mit dieser neuen Methode jederzeit bestimmt werden, wie hoch die Trefferrate oder wie niedrig die Fehlalarmrate sein müssen, damit sich ein wirtschaftlicher Nutzen ergibt. Obendrein wird dabei sogar das gesamte Spektrum an Störungsanteilen und Kosten-Nutzen-Verhältnissen abgedeckt. Durch die Visualisierung kann zudem die Evaluation auch graphisch erfolgen, wodurch die Stärken des originären, zweidimensionalen ROC-Graphs erhalten bleiben. Dies gilt mit Einschränkung sogar für den vierdimensionalen Ansatz. Insbesondere die Break Even Situation als Schwelle, ab der ein Klassifikator aus ökonomischer

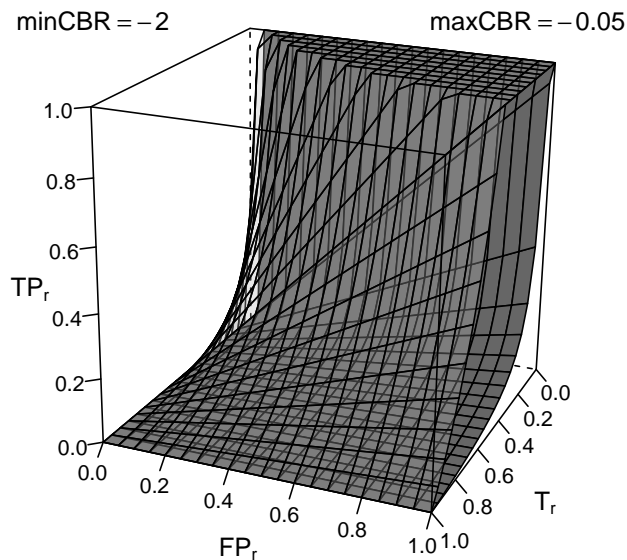


Abbildung 4-3: Beispiel für die Visualisierung der Break Even Situation bei Verwendung eines besten und eines schlechtesten Kosten-Nutzen-Verhältnisses (CBR)

Sicht ausreichend genau arbeitet, ist eine wichtige Eigenschaft bei der Evaluation. Denn an dieser Schwelle entscheidet sich, ob das Vorhersagemodell in der Lage ist, auch die Instandhaltungskosten zu senken und somit eines der Hauptziele der Arbeit zu erreichen. Um die entwickelte Methode im weiteren Verlauf der Arbeit anwenden zu können, ist allerdings noch zu klären, aus welchen Positionen sich die Kosten und der Nutzen zusammensetzen. D.h. es ist zu bestimmen, wie sich die Werte in der Kostenmatrix zusammensetzen.

4.2.3 Zusammensetzung der Kosten von Fehlalarmen und des Nutzens von verhinderten Störungen

Der vorangegangene Abschnitt stellte eine eigens entwickelte Methode vor, mit der die Genauigkeit einer Vorhersage aus wirtschaftlicher Sicht evaluiert und visualisiert werden kann (insbesondere die Break Even Situation). Er ließ aber offen, wie die Werte für Kosten und Nutzen im Detail ermittelt werden können. Diese Werte werden dazu benötigt, im konkreten Fall nachzuweisen, dass das zu entwickelnde Prognosemodell nicht nur Störungen verhindern kann, sondern auch die Instandhaltungskosten senkt.

Ausgangspunkt der Betrachtung ist die Kostenmatrix \mathbf{K} eines binären Klassifikationsproblems, wie sie in Gleichung (4.9) eingeführt wurde.

$$\mathbf{K} = \begin{bmatrix} K_{TP} & K_{FN} \\ K_{FP} & K_{TN} \end{bmatrix}$$

Im folgenden wird aufgeschlüsselt, was sich in ihren Zellen verbirgt:

- Bei TP wird der Weichenzustand richtigerweise als gestört klassifiziert. Es wird also korrekt prognostiziert, dass die Weiche in weniger als fünf Tagen ausfällt. Daraufhin werden Instandhaltungsmaßnahmen ausgeführt, die den Zustand der Weiche soweit verbessern, dass es nicht zu einer Störung kommt. Dies erfolgt so rechtzeitig, dass der Bahnbetrieb nicht außerplanmäßig beeinträchtigt wird, weil Betriebspausen für die Ausführung genutzt

werden. Aus der Verwendung des Prognosemodells entsteht somit ein Nutzen in Höhe der eingesparten Kosten bei einer Störung. Es ist zu beachten, dass die Kosten der eigentlichen Instandhaltungsmaßnahme nicht zu berücksichtigen sind, da diese auch entstünden, wenn es kein Prognosemodell gäbe oder die Störung nicht rechtzeitig klassifiziert würde. Die direkten Kosten der Instandhaltung (vgl. Abschnitt 2.2.3) können somit nicht eingespart werden.

$$K_{TP} = K_{FN} = B_{TP} \quad (4.19)$$

- FN ist eine Störung die auftritt, weil das Prognosemodell sie nicht erkannt hat. Bei der Entstörung entstehen direkte Instandhaltungskosten, die jedoch analog zum Nutzen von TP nicht zu berücksichtigen sind. Als Folge der Nichtverfügbarkeit entstehen indirekte Instandhaltungskosten. Im Eisenbahnsektor hat sich der Begriff der Betriebserschwerungskosten (BEK) als Synonym für die indirekten Kosten etabliert. Diese setzen sich in Anlehnung an Veit und Petri [VP08] zusammen aus:
 - den Opportunitätskosten durch entgangene Einnahmen bzw. Gewinne,
 - den Kosten als Folge von Pönalen,
 - den Kosten aus verpassten Anschlusszügen,
 - den Kosten für Ersatzverkehr,
 - den zusätzlichen Betriebsführungskosten,
 - den Kosten aus negativen Marktreaktionen und
 - den Umleitungskosten.

Zur Vereinfachung werden BEK über die entstandenen V_{Min} und einen Kostensatz je V_{Min} abgerechnet. Letzterer bildet eine Art Durchschnitt der einzelnen Positionen der BEK, der z.B. in den Arbeiten von Schilling und Lücking [SL03], Schönemann [Sch09], Gutsche [Gut10] und von vielen Eisenbahninfrastrukturunternehmen (EIU) verwendet wird, wenn auch in unterschiedlicher Höhe. Im weiteren Verlauf der Arbeit werden die BEK daher ebenfalls als ein entsprechender Kostensatz je V_{Min} (K_{VMin}) berücksichtigt.

$$K_{FN} = V_{Min} \cdot K_{VMin} \quad (4.20)$$

- Bei FP werden in Folge eines Fehlalarms unnötige Instandhaltungsmaßnahmen initiiert. In Anlehnung an die DIN EN 60300-3-3 [DIN03] würden sich die Kosten für die Instandhaltung theoretisch aus den folgenden Positionen zusammensetzen:
 - den Fahrtkosten zur Weiche,
 - den Personalkosten zur Durchführung sowie für An- und Abrückzeit,
 - den Materialkosten und
 - den entgangenen Trassenerlösen während der Streckensperrung zur Instandhaltung.

Praktisch betrachtet sind jedoch nur einige Positionen relevant. Auf Basis der Information aus der Zustandsprognose wird eine Störung erwartet und eine entsprechende Maßnahme geplant. Daraufhin begibt sich das Instandhaltungspersonal zur betreffenden Weiche an

die Stecke. Dort stellen die Mechaniker nach kurzer Überprüfung allerdings fest, dass die Weiche funktionsfähig ist. Ohne Materialeinsatz begeben sie sich zurück zur Zentrale oder zum nächsten Einsatz. Weil die Maßnahme rechtzeitig geplant werden konnte, wird sie während der Betriebspausen durchgeführt und mindert daher auch nicht die Trassenerlöse.

$$\begin{aligned}
 K_{FP} &= K_{Fa} + K_{Pe} \\
 &= \text{Kraftstoffpreis} \cdot \text{Verbrauch} \cdot (\text{Anfahrstrecke} + \text{Abfahrstrecke}) \\
 &\quad + \text{Personalkostensatz} \cdot (\text{Wegezeit} + \text{Maßnahmendauer}) \cdot \text{Personen}
 \end{aligned} \tag{4.21}$$

- TN verursacht keine Kosten. Die Weiche wird als noch mindestens fünf Tage funktionierend klassifiziert, was auch tatsächlich so ist. Es sind also keine Maßnahmen durchzuführen.

$$K_{TN} = 0 \tag{4.22}$$

Mit den Gleichungen (4.19), (4.20), (4.21) und (4.22) wird es nun möglich, die einzelnen Werte der Kostenmatrix **K** zu bestimmen, um wiederum im konkreten Fall die Genauigkeit eines Vorhersagemodells aus wirtschaftlicher Sicht bewerten zu können. In Verbindung mit der vorgestellten Evaluationsmethode kann somit nachgewiesen werden, dass das zu entwickelnde Prognosemodell dazu beiträgt, die Instandhaltungskosten von Weichen zu senken.

Nummer	Kurzbeschreibung	Absicht
1.	Vorhersage in Form einer RUL	Planbarkeit der Instandhaltung
2.	Vorhersage mit TP Rate größer 0.5 für die Klasse der RUL von fünf oder weniger Tagen	mehr Pünktlichkeit in Deutschland
3.	Summe der VMin von TP in der Klasse mit der RUL von fünf oder weniger Tagen ist größer 50%	mehr Pünktlichkeit in Deutschland
4.	Erreichen der Break Even Situation	Wirtschaftlichkeit der Prognose
5.	Einsatz von so wenig Sensoren wie möglich	Wirtschaftlichkeit der Prognose
6.	Nutzung von Sensoren, die für den regulären Bahnbetrieb zugelassen sind	Wirtschaftlichkeit der Prognose
7.	Verwendung von Weichen mit verschiedenen Eigenschaften	Allgemeingültigkeit der Ergebnisse
8.	Verwendung von mindestens zehn Weichen	Allgemeingültigkeit der Ergebnisse
9.	Verwendung von häufig im Netz der Deutsche Bahn Aktiengesellschaft (DB AG) verbauten Weichen	Anwendbarkeit in Deutschland
10.	Erhebung von realen Mess- und Störungsdaten	Anwendbarkeit in Deutschland
11.	Datenbasis über den Zeitraum von mindestens einem Jahr	Anwendbarkeit in Deutschland

Tabelle 4-1: Auflistung aller Anforderungen an die zu entwickelnde Prognose

Alle in diesem Kapitel formulierten Anforderungen werden in der Tabelle 4-1 noch einmal zusammengefasst. Gemeinsam mit der entwickelten Methode zu ihrer Bewertung sollen sie sicherstellen, dass die Prognose effektiv, praktikabel und wirtschaftlich ist und demzufolge die folgenden aus den Forschungsfragen formulierten Ziele der vorliegenden Arbeit erreicht werden:

1. Mit dem Verfahren lassen sich mindestens **50 Prozent der Weichenstörungen rechtzeitig vorhersagen**. Das bedeutet, dass Störungen mit einer TP Rate von mindestens 0.5 in die

- Klasse mit einer RUL von fünf oder weniger Tagen sortiert werden, da sie so durch geplante Instandhaltungsmaßnahmen verhindert werden können.
2. Durch das Verfahren wird der Netzbetreiber in die Lage versetzt, auch **50 Prozent der durch Weichen verursachten Verspätungsminuten zu reduzieren**. Um dieses Ziel zu erreichen, müssen die TP in der entsprechenden Klasse für mindestens die Hälfte der VMin verantwortlich sein. Dabei muss diese Prädiktion zusätzlich bei zehn häufig vorkommenden Weichen mit unterschiedlichen Eigenschaften funktionieren, deren Überwachungsdaten aus dem Feld stammen und mindestens einen Zeitraum von einem Jahr umfassen.
 3. Das Verfahren arbeitet wirtschaftlich, d.h. die Lösung trägt auch zur **Senkung der Instandhaltungskosten** bei. Das ist der Fall, wenn die Vorhersagegenauigkeit die Break Even Situation erreicht. Außerdem kommen nur zugelassene Sensoren zur Weichenüberwachung zum Einsatz, damit keine neuen Zulassungskosten entstehen.
 4. Die Überwachungsdaten, auf denen die Prognose beruht, werden mit einer **minimalen Sensorausrüstung** gemessen, die **vollständig rückwirkungsfrei auf die Leit- und Sicherungstechnik** ist.

4.3 Zusammenfassung des Kapitels

Aus den Zielen der vorliegenden Arbeit und der Kritik an existierenden Modellen zur Zustandssprognose für Weichen wurden in diesem Kapitel Anforderungen abgeleitet, die an das zu entwickelnde Prognosemodell gestellt werden. Neben den allgemeinen Anforderungen wurde vor allem auf die Wirtschaftlichkeit der Vorhersage Wert gelegt. Wird ein Prognosemodell angewendet, bestimmt sich dessen Wirtschaftlichkeit durch die Genauigkeit. Es müssen ausreichend viele Störungen richtig und rechtzeitig erkannt werden, während gleichzeitig eine bestimmte Zahl an Fehlalarmen nicht überschritten werden darf.

Um diese Genauigkeitsanforderung aus wirtschaftlicher Sicht bestimmen und messen zu können, wurde eigens eine neue Methode entwickelt. Dazu wurde zunächst die stetige verbleibende Nutzungsdauer (engl. Remaining Useful Life) (RUL) in ein binäres Klassifikationsproblem transformiert, wobei erstmals auch die Frage der rechtzeitigen Offenbarung einer Störung beantwortet wurde. Rechtzeitig bedeutet hier fünf Tage vor dem eigentlichen Störungseintritt, damit entsprechende Instandhaltungsmaßnahmen geplant und während einer Betriebspause ausgeführt werden können. Die entwickelte Bewertungsmethode erweitert die zwei Dimensionen des Receiver Operating Characteristic (ROC)-Graphs aus falsch Positiv (FP) Rate und richtig Positiv (TP) Rate um die Dimensionen des Störungsanteils an allen Messungen und des Kosten-Nutzen-Verhältnisses von FP und TP. Damit kann die Genauigkeit nicht nur über das gesamte Spektrum an Störungshäufigkeiten und unterschiedlichen Kosten bzw. Nutzen bewertet werden. Sondern es kann insbesondere die Break Even Situation der Vorhersage bestimmt werden, ab der die Zustandsprognose die Instandhaltungskosten senkt. Vor allem ihre Visualisierung im dreidimensionalen Raum ermöglicht die schnelle und einfache Bewertung vom ökonomischen Standpunkt. Abschließend wurde detailliert, aus welchen Positionen sich die Kosten eines Fehlalarms und der Nutzen einer richtigen Vorhersage zusammensetzen, um letztlich die Bewertungsmethode im konkreten Fall – z.B. im weiteren Verlauf der Arbeit – anwenden zu können.

5 Datenquellen und Datenbasis für die Entwicklung des Prognosemodells

Wesentliche Grundlagen zum Aufbau und zur Funktion von Weichen sowie zur zustandsorientierten Instandhaltung wurden in Kapitel 2 beschrieben. In Kapitel 3 standen dann allgemein Verfahren der Datenanalyse und der Zustandsprädiktion im Fokus, worauf Kapitel 4 Anforderungen formulierte, die an das zu entwickelnde Prognosemodell gestellt werden, damit dieses effektiv, praktikabel und wirtschaftlich ist. Das folgende Kapitel beschäftigt sich mit der Selektion und Vorverarbeitung von Daten, die als Basis für das Prognosemodell unverzichtbar sind. Es gilt, jene Parameter zu erfassen, die Aufschluss über den Weichenzustand geben und gleichzeitig die definierten Anforderungen erfüllen. Dazu werden zunächst Daten eines kommerziellen Weichendiagnosesystems analysiert (Abschnitt 5.1), um deren Eignung für die Prognose zu bestimmen. Anschließend folgt eine systematische Analyse, welche Daten gegebenenfalls zum Zweck der Prognose zu ergänzen sind (Abschnitt 5.2). Dieser Teil ist insbesondere deswegen wichtig, weil ein derart strukturiertes Vorgehen in der Fachwelt bisher fehlt. Als Vorbereitung der Prognosemodellentwicklung wird abschließend die relevante, verfügbare Datenbasis vorgestellt (Abschnitt 5.3).

Mit diesem Kapitel folgt die Arbeit dem Prozess der Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD), dessen erste Schritte die Datenselektion und die Datenvorverarbeitung sind. Wie in Abschnitt 3.1.2 hinlänglich argumentiert wurde, ist der Prozess der KDD bestens geeignet, um sich daran bei der Entwicklung einer datengetriebenen Zustandsvorhersage zu orientieren.

5.1 Analyse von Messdaten eines Weichendiagnosesystems und ihrer Eignung zur Zustandsbeurteilung

Für die vorliegende Arbeit sind Daten das Fundament zur Zustandsbeurteilung. In Anbetracht der zahlreichen Weichenbestandteile und deren vielfältigen Ausprägungen (vgl. Abschnitt 2.1.2) wird deutlich, dass es sehr schwer ist, alle Parameter zu erfassen, die direkt Aufschluss über den Weichenzustand geben. Was für die Oberbaukomponenten (vgl. Abbildung 2-1) noch realisierbar sein mag, ist für die Komponenten der Eisenbahnleit- und -sicherungstechnik (LST) vorerst unmöglich. Doch ein Großteil der Weichenstörungen geht nun mal von den LST-Komponenten aus [JZ02, Zwa09, Böh13c, RA13]. Gemäß des Vorschlags von Böhm und Gutsche [BG11] zur Umsetzung der zustandsorientierten Instandhaltung sollen daher Parameter gemessen werden, die indirekt über den Zustand Aufschluss geben (siehe Abschnitt 2.2.5). Im Folgenden wird die Entscheidung für ein Weichendiagnosesystem als Quelle solcher Parameter begründet (Abschnitt 5.1.1) und der Datenumfang näher beschrieben (Abschnitt 5.1.2). Außerdem werden Störungsdaten als Referenz für den tatsächlichen Zustand eingeführt (Abschnitt 5.1.3), damit anschließend die technische (Abschnitt 5.1.4) und die wirtschaftliche (Abschnitt 5.1.5) Performance des Weichendiagnosesystems evaluiert werden kann. Insgesamt wird so untersucht, inwieweit eine Prognose anhand von Daten des Weichendiagnosesystems möglich ist.

5.1.1 Entscheidung für SIDIS W als Weichendiagnosesystem

Parameter, die indirekt Aufschluss über den Zustand geben, sind z.B. die elektrische Stromstärke (I), die elektrische Spannung (U) und die Leistung (P) des Antriebs, aber auch die Kraft (F) während der Umstellung und die Dauer des Umstellvorgangs. Die Stellkraft ist zwar am aussagekräftigsten für die Zustandsbeurteilung, aber im Grunde kann mit jedem Parameter der Zustand der LST-Komponenten interpretiert werden. So erklärt sich auch der hohe Anteil an bestehenden Diagnose- und Prognoseansätzen, die diese Parameter verwenden (vgl. Tabelle 2-2).

Auch die beiden vorgestellten Weichendiagnosesysteme SIDIS W (compact) von Siemens und Roadmaster von voestalpin nutzen zum Teil diese Größen. Gegenüber vielen experimentell eingesetzten Sensorausrüstungen, sind beide Systeme durch das Eisenbahn-Bundesamt (EBA) für den Bahnbetrieb in Deutschland zugelassen (siehe Abschnitt 2.3.2 und Tabelle 2-1 zu allgemeinen Systemeigenschaften). Sie stellen dadurch nicht nur den Stand der Technik dar, sondern erfüllen vor allem die entsprechende Anforderung an das zu entwickelnde Prognosemodell (siehe Nummer 6 in Tabelle 4-1).

Im Rahmen der Forschung zur vorliegenden Arbeit wurden im Jahr 2009 Daten akquiriert. Damals wurde SIDIS W aus folgenden Gründen gegenüber Roadmaster bevorzugt:

1. SIDIS W war zu dem Zeitpunkt bei der DB Netz AG weiter verbreitet und länger in der Anwendung.
2. Das System lieferte mehr Parameter zur Interpretation und wurde vom Betreiber als das bessere angesehen [Gut08].
3. Der Betreiber konnte den Datenexport per Fernzugriff aus der Zentrale durchführen. Bei Roadmaster hätten die XML-Dateien vor Ort auf ein physisches Medium kopiert werden müssen, weil die Pilotanlage nicht mit dem Kommunikationsnetzwerk verbunden war.
4. Da die Weichenmotoren mit Dreiphasenwechselstrom betrieben werden, sorgt die Asymmetrie des Stroms dafür, dass die Wirkleistung des Antriebs eher die benötigte Umstellkraft spiegelt, als die Stromstärke. Weil SIDIS W mit der Wirkleistung arbeitet und somit die Asymmetrie berücksichtigt, korrelieren dessen Zustandsparameter stärker mit der Stellkraft als beim Roadmaster-System, das auf der Stromstärke basiert. Es wurde angenommen, dass die Wirkleistung besser für die Zustandsbeurteilung geeignet ist als die Stromstärke. Diese Annahme wurde später durch andere wissenschaftliche Arbeiten bestätigt (vgl. [GMRT10, ARK13]).

Im Anfangsstadium der Forschung zur vorliegenden Arbeit konnte bereits nachgewiesen werden, dass die Messwerte im Langzeittrend eine Zustandsverschlechterung erkennen lassen (vgl. [GB11]). Es ist allerdings zu klären, ob das System in seiner bestehenden Form ausreichend genau arbeitet. Zuvor wird kurz der Umfang an akquirierten Daten vorgestellt und außerdem erläutert, welche Referenz für den tatsächlichen Weichenzustand genutzt wird.

5.1.2 Zeitintervall der akquirierten SIDIS W Messdaten

Die DB Netz AG überließ dem DLR Messdaten von **29 Weichen**, die an SIDIS W angeschlossen sind. Diese umfassen den Zeitraum vom **01. März 2007 bis zum 31. März 2009**. Der Betreiber

konnte zwar nicht für alle Weichen über den gesamten Zeitraum Daten bereitstellen, aber für 17 von ihnen ist die Anforderung erfüllt, dass der Zeitraum mindestens ein Jahr und somit alle Jahreszeiten umfasst (siehe Nummer 11 in Tabelle 4-1). Insgesamt ist die Menge von 294796 Messungen ($|S| = 294796$) zu Analysezwecken verfügbar.

Von welchen Weichen über welchen Zeitraum Daten vorliegen, veranschaulicht Abbildung 5-1. Bei der Betrachtung der Messhäufigkeit fällt auf, dass für die Weichen unterschiedlich viele Messungen am Tag aufgezeichnet wurden. SIDIS W misst nur, wenn die Weiche umläuft. Da einige Weichen betrieblich stärker beansprucht werden als andere, also häufiger umgestellt werden, liegen dafür auch mehr Messungen vor. Die Weiche 29 wird so selten benutzt, dass an manchen Tagen sogar überhaupt keine Aufzeichnung erfolgte. Weiche 21 ist hingegen so stark frequentiert, dass sie mehr als 100 Mal am Tag umläuft. Abgesehen von diesen Schwankungen sind Aufzeichnungslücken festzustellen, bspw. im November 2007 und 2008 bei den Weichen 1 bis 15. Sie liegen alle in einem größeren Bahnhof. Bei der jährlichen Instandhaltung des Fahrweges wurde die Aufzeichnung für einige Tage ausgesetzt. Bei größeren Aufzeichnungslücken von mehr als einem Monat konnte der Betreiber nicht mehr auf die manuell gesicherten Daten zugreifen. Nichtsdestotrotz sind die Daten wertvoll, da sie von Weichen im realen Betrieb stammen, womit auch diese Anforderung erfüllt wird (siehe Nummer 10 in Tabelle 4-1).

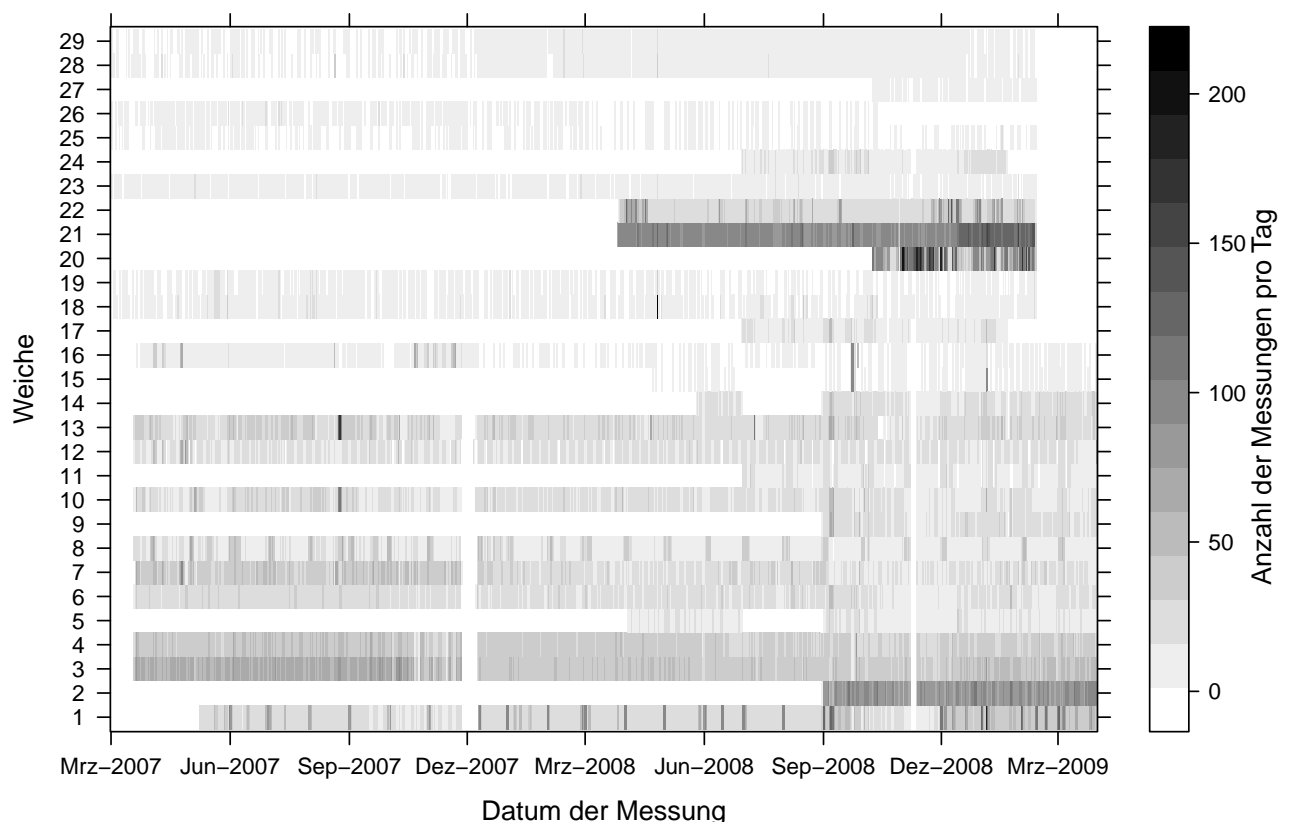


Abbildung 5-1: Weichen und Zeitraum verfügbarer Messdaten aus SIDIS W

Messdaten allein sind jedoch noch nicht ausreichend. Um evaluieren zu können, wie gut sich SIDIS W in seiner originären Form für die Zustandsbeurteilung eignet, braucht es vor allem Daten über den tatsächlichen Zustand der Weichen.

5.1.3 Störungsdaten als Referenz für den Weichenzustand

Bisher wurden die Entscheidung für SIDIS W als Datenquelle begründet und der Zeitraum akquirierter Messdaten veranschaulicht. Daten über den tatsächlichen Weichenzustand sind als Referenz für die Zustandsprognose unverzichtbar. Auch für die Evaluation von SIDIS W wird solch eine Referenz benötigt. Da es keine kontinuierliche, direkte Information über den Zustand gibt – sonst bräuhete kein Modell entwickelt werden – sind Störungen die einzige verlässliche Quelle zum Zustand. Sie bezeichnen den Zustand der Anlage zu einem ganz bestimmten und seltenen Zeitpunkt, nämlich wenn der Abnutzungsvorrat (vgl. Abschnitt 2.2.1) aufgebraucht und die verbleibende Nutzungsdauer (engl. Remaining Useful Life) (RUL) null ist.

Die DB Netz AG dokumentiert ihre Störungen in der betrieblichen Unternehmenssoftware SAP. Ein Auszug dieser Daten zu 575 Weichenstörungen der betreffenden Streckenabschnitte wurde dem DLR ebenfalls zur Analyse überlassen. Er umfasst alle Störungen im Gesamtzeitraum, für den auch die Daten aus SIDIS W vorliegen, also von März 2007 bis März 2009. Allerdings sind darin rund 80 weitere Weichen enthalten, die nicht an ein Diagnosesystem angeschlossen sind. Werden diese Störungen abgezogen ebenso wie Störungen, denen wegen Aufzeichnungslücken keine SIDIS W Daten gegenüberstehen, so verbleiben **98 Störungen**.

Die SAP-Einträge zu Störungen liefern einige Attribute, die für die Interpretation der Störung und für die Modellentwicklung interessant sind. Die wichtigsten sind in der Tabelle 5-1 aufgelistet, während eine vollständige Übersicht in der Tabelle B-2 im Anhang B.2 aufgeführt ist. Beginn und Ende der Störung (StörBegDatum, StörEndDatum) sowie der technische Platz (TechnPlatz) sind für die richtige Zuordnung der Störung zu Messzeitpunkt und Weiche bei SIDIS W wichtig. Die übrigen Attribute (Beschreibung, Priorität, Ursachentext, ProblCodeTxt, ProblGrpText) ermöglichen es, zu entscheiden, ob eine Störung überhaupt hätte prognostiziert werden können. Wenn bspw. ein Fremdkörper zwischen Backenschiene und Zunge liegt und dadurch den Umlauf blockiert (mechanische Hemmung), ist dies in der Regel ein zufälliges, plötzliches Ereignis und nicht vorhersagbar.

Attribut	Bedeutung
StörBegDatum	Datum und Uhrzeit des Störungsbeginns
StörEndDatum	Datum und Uhrzeit des Störungsendes
TechnPlatz	eindeutige SAP-spezifische Identifikation der Anlage, inklusive Ort, Typ und Bauteilen
Beschreibung	Freitext zur Beschreibung der Störung
Priorität	Priorität, mit der die Störung behoben werden muss
Ursachentext	Freitext zur Beschreibung der Störungsursache
ProblCodeTxt	definierter Text für die Störungsursache
ProblGrpText	definierter Text für die gestörte Anlagengruppe

Tabelle 5-1: Ausgewählte Attribute des SAP-Auszugs von Störungsmeldungen und ihre Bedeutung

Diese Daten sind ebenfalls im realen Betrieb erhoben und erfüllen somit die gleichlautende Anforderung an die zu entwickelnde Prognose. Werden beide Quellen (Mess- und Störungsdaten) einander gegenübergestellt, lässt sich die Performance von SIDIS W systematisch evaluieren.

5.1.4 Performance Evaluation des Weichendiagnosesystems SIDIS W unter technischen Gesichtspunkten

SIDIS W misst die Wirkleistung des Antriebs, während die Weiche umgestellt wird. Wie im Abschnitt 2.3.2 bereits erwähnt wurde, ergibt sich in der Regel eine typische Umlaufkurve. Zu einem frei wählbaren Zeitpunkt wird für jede Weiche die Leistungskurve gemessen und als Referenzmessung festgelegt. Die Störungsdetektion erfolgt anhand konfigurierbarer Grenzwerte für die Abweichung zur Referenzmessung. Bei Über- oder Unterschreiten eines ersten Grenzwertes wird ein gelber Alarm ausgelöst, wie in Abbildung 5-2 dargestellt. Steigt bzw. sinkt der Messwert weiter, wird ein roter Alarm ausgelöst. Der Hersteller Siemens empfiehlt bei Gelb eine baldige Wartung, während bei Rot eine sofortige Wartung notwendig ist (vgl. [KR11, S. 7]).

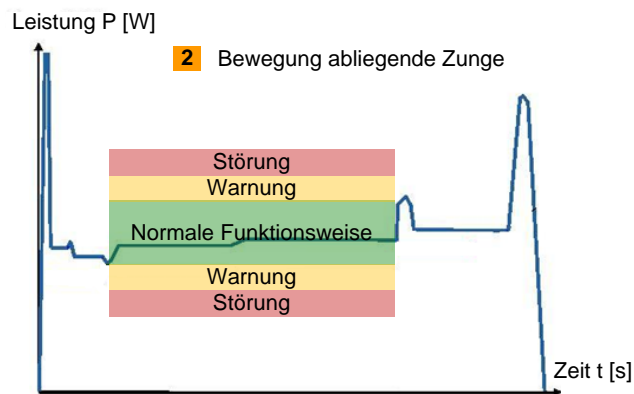


Abbildung 5-2: Prinzip der Alarme in SIDIS W am Beispiel der idealisierten Wirkleistungskurve (in Anlehnung an [SB02, Bol07])

Dass dieses Prinzip der Grenzwerte und Alarme sich nicht gut zur Zustandsbeurteilung eignet, ist aus Veröffentlichungen im Rahmen der Forschung zur vorliegenden Arbeit bekannt (vgl. [Böh12a, Böh12b, Böh12c]). Alarme werden zu häufig ausgelöst, als dass sie tatsächlich eine Störung bedeuten oder gar ankündigen. Im Fall der vorliegenden SIDIS W Daten ist die Alarmanzahl so hoch, dass die Instandhaltung sie schon bald nach der Einführung 2002 ignorierte [Gut08]. Für die Gegenüberstellung von tatsächlichen Störungen und Alarmen ist das allerdings ein Vorteil, weil die Störungen garantiert unabhängig von den Messungen sind. Da SIDIS W auch aufzeichnet, welcher Alarm ausgelöst wurde, kann die Anzahl der jeweiligen Alarme im Vergleich zu den tatsächlichen Störungen nachvollzogen werden. Damit gelingt eine objektive, messbare Evaluation des Systems, unabhängig von der subjektiven Akzeptanz der Nutzer.

Welche Diskrepanz zwischen Alarmen und tatsächlichen Störungen besteht, zeigt die Tabelle 5-2. Es treten nicht nur sehr viel mehr rote Alarme auf als Störungen, sondern es existiert auch keine Korrelation zwischen der Alarmanzahl und den Störungen. Bspw. war Weiche 8 verhältnismäßig häufig gestört, zeigt aber den geringsten Prozentsatz an roten Alarmen aller Weichen. Bei Weiche 28 wurde nahezu in jedem dritten Umlauf ein roter Alarm ausgelöst, eine Störung ist aber nicht eingetreten. Angesichts solcher Zahlen ist auch die in Abschnitt 2.3.2 zitierte Unzufriedenheit der DB Netz AG mit den Diagnosesystemen verständlich, die Holst [Hol12, S. 20] in seinem Artikel zum Ausdruck bringt.

Um die Genauigkeit von SIDIS W zu evaluieren, wird die in Abschnitt 3.3 beschriebene Receiver Operating Characteristic (ROC)-Graph Metrik verwendet. Ausgangspunkt ist die Konfusions-

Weiche	Messung von	Messung bis	Anzahl Messungen	Anzahl gelber Alarme (ohne Rot)	% der Messungen	Anzahl roter Alarme	% der Messungen	Tatsächlich dokumentierte Störung
1	9. Mai. 07	31. Mrz. 09	22458	3794	16.9 %	2255	10.0 %	5
2	31. Aug. 08	31. Mrz. 09	17547	7519	42.9 %	3771	21.5 %	8
3	19. Mrz. 07	31. Mrz. 09	34887	4198	12.0 %	11276	32.3 %	4
4	19. Mrz. 07	31. Mrz. 09	26360	9600	36.4 %	2647	10.0 %	3
5	3. Apr. 08	31. Mrz. 09	4087	1364	33.4 %	357	8.7 %	1
6	19. Mrz. 07	31. Mrz. 09	15275	1738	11.4 %	247	1.6 %	2
7	20. Mrz. 07	31. Mrz. 09	19854	8782	44.2 %	3299	16.6 %	1
8	19. Mrz. 07	31. Mrz. 09	13499	800	5.9 %	178	1.3 %	10
9	31. Aug. 08	31. Mrz. 09	4500	1665	37.0 %	567	12.6 %	1
10	19. Mrz. 07	31. Mrz. 09	14974	6366	42.5 %	2012	13.4 %	4
11	1. Jul. 08	31. Mrz. 09	3220	1324	41.1 %	633	19.7 %	2
12	19. Mrz. 07	31. Mrz. 09	12198	3626	29.7 %	645	5.3 %	2
13	19. Mrz. 07	31. Mrz. 09	19718	3822	19.4 %	690	3.5 %	7
14	26. Mai. 08	31. Mrz. 09	4904	1221	24.9 %	1933	39.4 %	2
15	22. Apr. 08	31. Mrz. 09	804	331	41.2 %	161	20.0 %	9
16	20. Mrz. 07	31. Mrz. 09	3004	824	27.4 %	143	4.8 %	1
17	30. Jun. 08	21. Jan. 09	2908	1537	52.9 %	1215	41.8 %	2
18	3. Mrz. 07	12. Feb. 09	3500	1083	30.9 %	59	1.7 %	2
19	3. Mrz. 07	12. Feb. 09	2195	772	35.2 %	191	8.7 %	1
20	9. Okt. 08	12. Feb. 09	11423	7806	68.3 %	2646	23.2 %	4
21	26. Mrz. 08	12. Feb. 09	31996	4178	13.1 %	435	1.4 %	13
22	27. Mrz. 08	12. Feb. 09	11009	1958	17.8 %	813	7.4 %	11
23	1. Mrz. 07	12. Feb. 09	3388	586	17.3 %	308	9.1 %	3
24	30. Jun. 08	21. Jan. 09	3129	2095	67.0 %	116	3.7 %	0
25	1. Mrz. 07	12. Feb. 09	1265	509	40.2 %	88	7.0 %	0
26	1. Mrz. 07	18. Okt. 08	1566	62	4.0 %	105	6.7 %	0
27	9. Okt. 08	12. Feb. 09	451	143	31.7 %	27	6.0 %	0
28	1. Mrz. 07	12. Feb. 09	2508	1078	43.0 %	811	32.3 %	0
29	1. Mrz. 07	12. Feb. 09	2169	79	3.6 %	75	3.5 %	0
Ge-samt	1. Mrz. 07	31. Mrz. 09	294796	78860	26.8 %	37703	12.8 %	98

Tabelle 5-2: Gegenüberstellung von Alarmen in SIDIS W mit den tatsächlichen Störungen

matrix $\mathbf{C} = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} = \begin{pmatrix} 27 & 71 \\ 37676 & 257022 \end{pmatrix}$ für die binäre Klassifikation (vgl. Abbildung 3-7). Daraus werden die falsch Positiv Rate (FP_r) und die richtig Positiv Rate (TP_r) gemäß den Gleichungen (4.10) und (4.11) berechnet. Wird sich an der Empfehlung des Herstellers orientiert, ist jeder rote Alarm als positive Klassifikation anzusehen. Doch erst wenn darauf unmittelbar auch eine Störung registriert wurde, liegt der TP Fall vor. Andernfalls handelt es sich um einen Fehlalarm. So ergibt sich eine $FP_r = 0.128$ und eine $TP_r = 0.276$. Die Fehlalarmrate spiegelt im Wesentlichen den Anteil roter Alarme an allen Messungen wieder. Dass allerdings nicht einmal einem Drittel aller Störungen ein roter Alarm vorausging, ist unerwartet. Wäre der Zusammenhang zwischen Stellstrom und Zustand nicht durch zahlreiche unabhängige Quellen belegt, würde es den Grundansatz dieser Forschungsarbeit infrage stellen.

Natürlich kann entgegengehalten werden, dass nicht jede Messung einzeln bewertet werden muss. Es könnten ja auch die Alarme mehrerer Messungen zusammengefasst werden, bspw. fünf aufeinander folgende Umläufe. Erstmalig wurde ein solcher Ansatz im Rahmen der Forschung zur vorliegenden Arbeit angewendet (siehe auch [Böh12a, Böh12b]). Er wurde systematisch untersucht, indem für die positive Klassifikation als Störung die folgenden Maßstäbe angewendet wurden:

- Es wurde ein roter Alarm ausgelöst (originäre Funktion von SIDIS W).
- Es wurde ein roter oder gelber Alarm ausgelöst.
- Es wurde bei x aufeinanderfolgenden Messungen je ein roter Alarm registriert. Wobei x 5, 10, 20 oder 50 ist ($x \in \{5, 10, 20, 50\}$).
- Es wurde bei x aufeinanderfolgenden Messungen je ein gelber oder roter Alarm registriert.
- In y Stunden sind z Prozent rote Alarme aufgetreten. Wobei y für 48, 96, 144 und z für 25, 33, 50, 67, 75, 80 steht ($y \in \{48, 96, 144\}$ und $z \in \{25, 33, 50, 67, 75, 80\}$)
- In y Stunden sind z Prozent gelbe oder rote Alarme aufgetreten.

Aus den verschiedenen Maßstäben ergeben sich ganz unterschiedliche Konfusionsmatrizen und demzufolge verschiedene Werte der Metrik. Zu Gunsten der Übersichtlichkeit werden in der Tabelle 5-3 nur die Werte der Metrik aufgelistet, also FP und TP Rate. Darin zeigt sich, dass eine Verbesserung des einen Wertes fast immer auch mit der Verschlechterung des anderen einhergeht. Es gibt zwar Maßstäbe, die besser sind als die ursprüngliche Funktionalität von SIDIS W (z.B. 25% rote Alarme in 96 Stunden), aber wirklich verlässlich ist keiner davon. Wie der ROC-Graph in Abbildung 5-3 zeigt, heben sie sich nur schwach von der zufälligen Zustandsbeurteilung ab, bei der FP und TP Rate gleich sind. (Auf eine Beschriftung oder Legende wurde hier bewusst verzichtet, um die Nähe zur Zufälligkeit aller Maßstäbe zu unterstreichen. Der detaillierte ROC-Graph kann in in Abbildung B-1 des Anhang B.1 eingesehen werden.) Aus technischer Sicht ist diese Ungenauigkeit nicht akzeptabel. Ob sie jedoch ggf. aus wirtschaftlicher Sicht ausreicht, wird im Folgenden mit der in Abschnitt 4.2 entwickelten Methode untersucht.

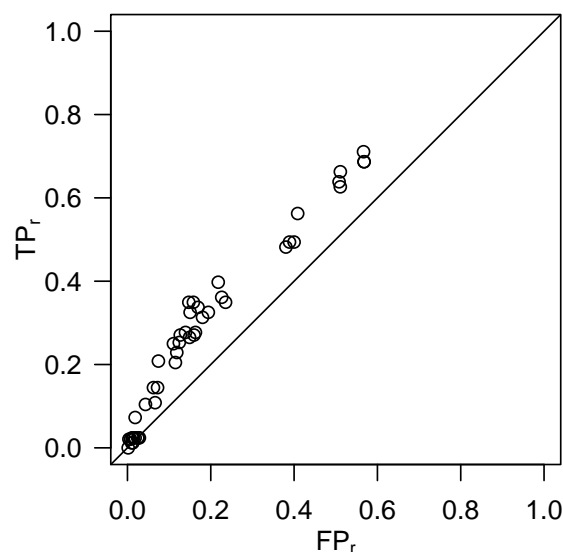


Abbildung 5-3: ROC-Graph der verschiedenen Maßstäbe für die Zusammenfassung mehrerer Alarme (zur Übersicht ohne Legende, ansonsten vgl. Abbildung B-1)

Zeilennr.	Maßstab	FP Rate	TP Rate
1	Einzelner roter Alarm (originäre Funktion von SIDIS W)	0.1268	0.2708
2	Einzelner gelber oder roter Alarm	0.4086	0.5625
3	5 aufeinander folgende rote Alarme	0.0185	0.0729
4	10 aufeinander folgende rote Alarme	0.0073	0.0208
5	20 aufeinander folgende rote Alarme	0.0035	0.0208
6	50 aufeinander folgende rote Alarme	0.0018	0.0000
7	5 aufeinander folgende gelbe oder rote Alarme	0.1599	0.2708
8	10 aufeinander folgende gelbe oder rote Alarme	0.1104	0.2500
9	20 aufeinander folgende gelbe oder rote Alarme	0.0743	0.2083
10	50 aufeinander folgende gelbe oder rote Alarme	0.0432	0.1042
11	25 % rote Alarme innerhalb von 48 Stunden	0.1583	0.3494
12	25 % rote Alarme innerhalb von 96 Stunden	0.1504	0.3253
13	25 % rote Alarme innerhalb von 144 Stunden	0.1473	0.3494
14	25 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.5667	0.7108
15	25 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.5682	0.6867
16	25 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.5677	0.6867
17	33 % rote Alarme innerhalb von 48 Stunden	0.1244	0.2530
18	33 % rote Alarme innerhalb von 96 Stunden	0.1184	0.2289
19	33 % rote Alarme innerhalb von 144 Stunden	0.1151	0.2048
20	33 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.5110	0.6627
21	33 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.5109	0.6265
22	33 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.5080	0.6386
23	50 % rote Alarme innerhalb von 48 Stunden	0.0723	0.1446
24	50 % rote Alarme innerhalb von 96 Stunden	0.0664	0.1084
25	50 % rote Alarme innerhalb von 144 Stunden	0.0624	0.1446
26	50 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.4000	0.4940
27	50 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.3890	0.4940
28	50 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.3804	0.4819
29	67 % rote Alarme innerhalb von 48 Stunden	0.0287	0.0241
30	67 % rote Alarme innerhalb von 96 Stunden	0.0257	0.0241
31	67 % rote Alarme innerhalb von 144 Stunden	0.0248	0.0241
32	67 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.2355	0.3494
33	67 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.2263	0.3614
34	67 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.2179	0.3976
35	75 % rote Alarme innerhalb von 48 Stunden	0.0184	0.0241
36	75 % rote Alarme innerhalb von 96 Stunden	0.0146	0.0241
37	75 % rote Alarme innerhalb von 144 Stunden	0.0129	0.0120
38	75 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.1941	0.3253
39	75 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.1802	0.3133
40	75 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.1694	0.3373
41	80 % rote Alarme innerhalb von 48 Stunden	0.0136	0.0241
42	80 % rote Alarme innerhalb von 96 Stunden	0.0113	0.0241
43	80 % rote Alarme innerhalb von 144 Stunden	0.0099	0.0120
44	80 % gelbe oder rote Alarme innerhalb von 48 Stunden	0.1632	0.2771
45	80 % gelbe oder rote Alarme innerhalb von 96 Stunden	0.1491	0.2651
46	80 % gelbe oder rote Alarme innerhalb von 144 Stunden	0.1393	0.2771

Tabelle 5-3: Werte der FP und TP Rate bei verschiedenen Maßstäben für die Zusammenfassung mehrerer Alarme

5.1.5 Performance Evaluation des Weichendiagnosesystems SIDIS W unter ökonomischen Gesichtspunkten

Nachdem im vorigen Abschnitt allgemein die Performance der SIDIS W Alarmer und ihrer systematischen Kombination bewertet wurde, folgt hier die Evaluation aus wirtschaftlicher Sicht. Dafür wird die in Abschnitt 4.2 entwickelte Methode verwendet, für die zunächst die Kosten eines Fehlalarms (K_{FP}) und der Nutzen einer richtig detektierten Störungen (B_{TP}) zu ermitteln sind. Dabei wird, wenn möglich, jeweils ein Minimum und ein Maximum der jeweiligen Kosten bzw. des Nutzens aufgestellt. Auf diese Weise kann ein schlechtestes und ein bestmögliches Kosten-Nutzen-Verhältnis (engl. Cost-Benefit-Ratio) (CBR) errechnet und anschließend für die Visualisierung genutzt werden.

Weil aus dem Jahr 2010 der Personalkostensatz für Instandhalter bekannt ist, werden auch alle anderen Kostensätze auf 2010 bezogen. Gemäß der Gleichung (4.21) ergeben sich die Kosten eines Fehlalarms aus den Fahrtkosten (K_{Fa}) und den Personalkosten (K_{Pe}), die durch die unnötig ausgelöste Wartungsmaßnahme entstehen (siehe Abschnitt 4.2.3). Im Einzelnen setzen sie sich aus den folgenden Positionen zusammen:

- Die Dieselpkraftstoffpreise lagen 2010 im Minimum der Monatsdurchschnitte bei 1.139 € und im Maximum bei 1.306 €. [ADA14]
- Um zum Einsatzort zu gelangen wird ein Dienstwagen verwendet. Das entsprechende Fahrzeug verbraucht auf 100 km 4.5 bis 6.0 Liter Diesel [Wik14].
- Die Weichen liegen in den Bahnhöfen Berlin Zoologischer Garten (BZOO) und Berlin Charlottenburg (BCHB), während das Instandhaltungspersonal am S-Bahnhof Bellevue stationiert ist. Je nach gewählter Route liegt BZOO 2.8 bis 3.4 km entfernt. Der Weg nach BCHB beträgt 3.5 bis 4.1 km. [Goo14]
- Der Personalkostensatz eines Instandhalters wurde mit 45 € je Stunde kalkuliert.
- Die Fahrzeit nach BZOO wird auf sechs bis zehn Minuten geschätzt, nach BCHB auf sechs bis elf, je nach Verkehrslage [Goo14]. Hinzu kommt noch ein Fußweg von einer bis fünf Minuten, um direkt zur Weiche zu gelangen. So addiert sich die Wegezeit für Hin- und Rückweg auf 14 bis 32 Minuten.
- Um festzustellen, dass die Weiche in Ordnung ist, benötigen die Instandhalter bestenfalls keine Zeit und schlimmstenfalls 15 Minuten.
- Ein Einsatz wird grundsätzlich von zwei Instandhaltern durchgeführt.

Eine richtig vorhergesagte Störung erspart genau jene Kosten, die durch die Störung entstanden wären. Der Nutzen von TP ergibt sich also, gemäß der Gleichung (4.20), aus den Verspätungsminuten (VMin) multipliziert mit dem Kostensatz je VMin (K_{VMin}). Für beide Positionen wurden folgenden Werte ermittelt:

- In den verfügbaren Störungsdaten sind keine Angaben zu den jeweiligen VMin enthalten, da die DB Netz AG diese Information bei der ursprünglichen Datenakquise nicht bereitgestellt hat. Eine spätere Rekonstruktion der Störungsauswirkungen war nicht mehr möglich. Um dennoch eine realistische Einschätzung zu den VMin je Störung zu erhalten, ermittelte die

DB Netz AG die durchschnittlichen VMin der betreffenden 29 Weichen aus den Jahren 2010 bis 2013. Bei 296 Weichenstörungen entstanden im arithmetischen Mittel 47.3 VMin. Es ist zwar aus Gesprächen mit Instandhaltern bekannt, dass die Verspätungswirkungen von Störungen stark schwanken können, je nach Verkehrsdichte und betrieblicher Relevanz der gestörten Weiche, aber für die Performanceanalyse besitzt der Durchschnittswert aus vier Jahren eine ausreichende Allgemeingültigkeit.

- Was eine Verspätungsminute kostet, gibt die Bahn nicht offiziell bekannt – auch nicht in Teilen und nicht einmal gegenüber der Bundesregierung, wie aus der nahezu unbeantworteten kleinen Anfrage im Bundestag hervorgeht [BfV10]. Inoffiziell kursieren sehr unterschiedliche Werte, die aber alle nicht belastbar sind oder nicht zitiert werden können. Deshalb werden Zahlen verwendet, deren Quellen allgemein zugänglich sind, auch wenn diese nicht von der DB Netz AG veröffentlicht wurden oder aus dem Ausland stammen.

Der niedrigste Wert für die Kosten je VMin beträgt 15 Euro und wird in den Dissertationen von Hempe [Hem06, S. 140] und Gutsche [Gut10, S. 118] verwendet. Beide Arbeiten beziehen sich auf Deutschland, ebenso wie der Artikel von Schilling und Lücking [SL03, S. 60, 68]. Darin verwenden die Autoren Betriebserschwerungskosten (BEK) von 60 bis 130 Euro je VMin.

Andere Quellen benennen nur einen Teil der BEK, vorwiegend die Pönalen. In ihrer Publikation verwenden García Márques et al. [GMLTR08, S. 1175] Durchschnittskosten, mit denen der britische Netzbetreiber Network Rail die betroffenen Eisenbahnverkehrsunternehmen (EVU) bei netzbedingten Zugverspätungen entschädigt. Die Kosten belaufen sich, abhängig von der Streckenkategorie, auf 40 bis 100 Britische Pfund je VMin. Dies entspricht rund 58 bis 146 Euro (umgerechnet mit dem durchschnittlichen Jahreswechsellkurs laut Bundesministerium der Finanzen [Bun08]). Bai schreibt in seiner Dissertation von bis zu 120 Britischen Pfund je VMin, die 2010 von Network Rail als Pönalen zu zahlen waren [Bai10, S. 4], rund 140 Euro (analog umgerechnet nach [Bun11]).

Im Gegensatz zu den Kosten von FP werden diese Werte nicht auf das Jahr 2010 angepasst. Weil BEK nicht Teil eines Warenkorbs sind und auch ihre Entwicklung über einen bestimmten Zeitraum nicht nachvollzogen werden kann, ist eine Anpassung an Zinsen oder Inflation nicht sinnvoll. Abgesehen davon führen die weite Spanne von 15 bis 146 Euro und die wenigen Zahlen dazwischen ohnehin zu einer eher unsicheren Einschätzung. Es ist daher umso wichtiger Minimum und Maximum zu verwenden.

Die einzelnen Kosten- und Nutzenpositionen sowie deren Minimum und Maximum sind in der Tabelle 5-4 noch einmal zusammengefasst. Sie werden für die Berechnung der CBR gemäß den Gleichungen (4.17) und (4.18) genutzt, mit denen wiederum die Break Even Situation visualisiert wird. Auf diese Weise wird nachgewiesen, ob SIDIS W oder die Kombination von Alarmen in jeder möglichen Situation, nur in bestimmten Fällen oder gar nicht wirtschaftlich arbeitet.

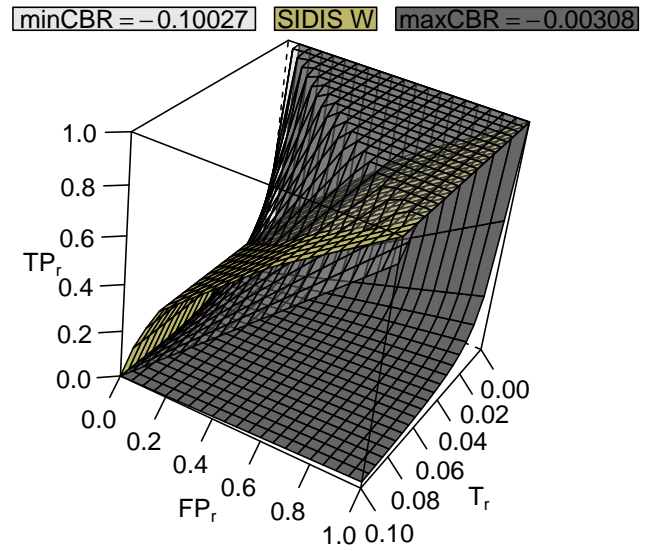
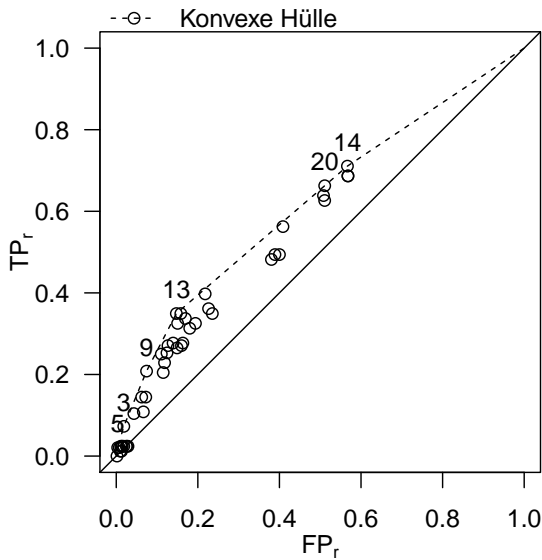
Dazu muss zunächst der zweidimensionale ROC-Graph mit der Performance der verschiedenen Maßstäbe der SIDIS W Alarme (siehe Abbildung 5-3) als Fläche in den dreidimensionalen Raum der Break Even Situation erweitert werden. Um aus dieser Ansammlung von Punkten eine Linie abzuleiten, wird die konvexe Hülle der SIDIS W Performance gebildet. Einfach ausgedrückt ist die konvexe Hülle die Verbindung von Punkten im ROC-Graph, die alle anderen Punkte dominieren

Position	Minimum	Maximum
Kraftstoffpreis [€/Liter]	1.139	1.306
Kraftstoffverbrauch [Liter/100 km]	4.5	6.0
Strecke für An- und Abfahrt [km]	5.6	8.2
Personalkostensatz [€/Stunde]	45	45
Wegezeit [Minuten (min)]	14	32
Maßnahmendauer [min]	0	15
Personen	2	2
Kosten FP = Kraftstoffpreis · Verbrauch · (Anfahrstrecke + Abfahrstrecke) + Personalkostensatz · (Wegezeit + Maßnahmendauer) · Personen		
Kosten FP [€]	21.29	71.14
VMin je Störung [min]	47.30	47.30
Kostensatz je VMin [€]	15	146
Nutzen TP = VMin · Kostensatz je VMin		
Nutzen TP [€]	709.50	6905.80

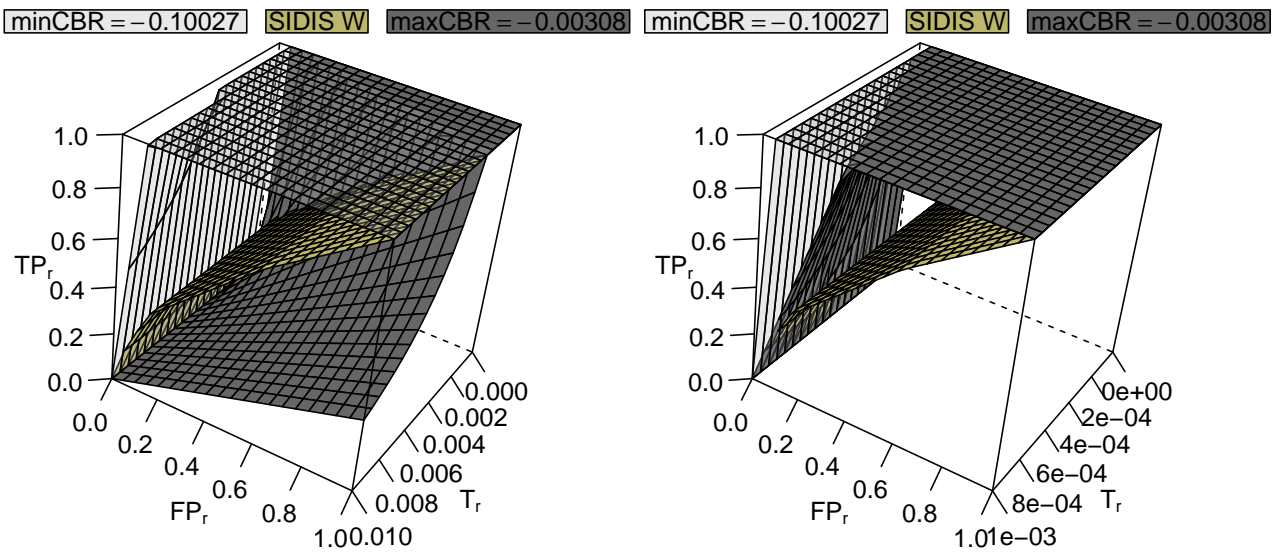
Tabelle 5-4: Positionen der Kosten eines Fehlalarms und des Nutzens einer richtig prognostizierten Störung

und somit die bestmögliche Performance aufzeigen [DH06, Han09]. In der Abbildung 5-4(a) ist die konvexe Hülle der Alarmkombinationen dargestellt (die Nummerierung der Punkte bezeichnet den entsprechenden Maßstab in der Tabelle 5-3). Diese konvexe Hülle wird wiederum parallel zur Achse des Störungsanteils (T_r) in die dritte Dimension verlängert, wodurch die Fläche der bestmöglichen SIDIS W Performance für die Break Even Situation entsteht. Zusammen mit den Flächen der minimalen und der maximalen CBR lässt sich nun aus den Abbildungen 5-4 die wirtschaftliche Performance ablesen.

Es zeigt sich in Abbildung 5-4(b), dass SIDIS W bis zu einem Störungsanteil von 0.09 einen Nutzen erzeugt, da SIDIS W bis dorthin die Fläche der Break Even Situation beim schlechtesten Verhältnis von Fehlalarmkosten zum Nutzen einer verhinderten Störung überdeckt. D.h. im Umkehrschluss wenn Weichen seltener als bei jedem 11. Umlauf gestört sind, erzeugt SIDIS W unter Umständen zusätzliche Kosten durch Fehlalarme – nämlich wenn z.B. die Instandhalter für den Weg zur Weiche und für die Überprüfung viel Zeit benötigen sowie die Kosten einer VMin mit 15 Euro veranschlagt werden. Bei einem Störungsanteil kleiner als 0.01 und größer als 0.001 könnte SIDIS W noch einen positiven Nutzen erzeugen, falls das Kostenverhältnis vorteilhaft ausfällt. Wie in Abbildung 5-4(c) zu sehen ist, liegt die konvexe Hülle des Weichendiagnosesystems hier zwischen den beiden Break Even Situationen. Allerdings sind im deutschen Netz kaum Weichen zu finden, die so häufig gestört sind. Auch bei 28 der analysierten 29 Weichen entstand die Störung erst nach mehr als 1000 Umlaufmessungen ($T_r \leq 0.001$) und bei 22 sogar erst nach mehr als 2000 Messungen ($T_r \leq 0.0005$). Bei 14 Weichen brauchte es sogar mehr als 4000 Messungen je Störung (vgl. Tabelle 5-2). Für SIDIS W bedeutet das, dass sein Einsatz mehr Kosten verursacht als einspart, sofern ein Anwender sich nach den Alarmen richtet. Das zeigt Abbildung 5-4(d) deutlich. Dass SIDIS W von den Instandhaltern nahezu vollständig ignoriert wird [Gut08, Böh13c], ist demzufolge nicht nur mit der mangelnden Akzeptanz bei den Nutzern begründbar, sondern auch aus ökonomischer Sicht sinnvoll. Aber wie ist diese mangelnde Performance zu begründen, wenn doch der Stellstrom ein Zustandsindikator ist? Antworten auf diese Frage liefert der kommende Abschnitt.



(a) Konvexe Hülle der verschiedenen Alarmmaßstäbe bei SIDIS W (Nummerierung aus Tabelle 5-3) (b) Auflösung wenn bei jedem zehnten Umlauf oder seltener eine Störung eintritt ($T_r \leq 0.1$)



(c) Auflösung wenn bei jedem 100. Umlauf oder seltener eine Störung eintritt ($T_r \leq 0.01$) (d) Auflösung wenn bei jedem 1000. Umlauf oder seltener eine Störung eintritt ($T_r \leq 0.001$)

Abbildung 5-4: Visualisierung der Break Even Situation von SIDIS W, aufgelöst nach verschiedenen Störungshäufigkeiten je Umlauf (T_r); bei positivem Nutzen liegt SIDIS W links bzw. oberhalb der anderen Flächen

5.2 Analyse weiterer Einflussfaktoren bei der Zustandsbeurteilung

Auf der einen Seite gelten aus dem Stellstrom abgeleitete Größen allgemein als Zustandsindikator, aber auf der anderen Seite hat der vorherige Abschnitt deutlich gezeigt, dass sich aus der alleinigen Schwellwertbetrachtung des Stellstroms keine ausreichend genaue Zustandsausage formulieren lässt. Wenn die Schwellwertverletzungen bzw. die daraus resultierenden Alarme geschickt kombiniert werden, verbessert sich die Performance zwar, bleibt aber dennoch zu ungenau. Dafür können im Wesentlichen zwei Gründe vorliegen:

1. Die Daten enthalten nicht die richtigen Größen, d.h. es fehlen eventuell Größen, die ebenfalls einen Einfluss auf den Zustand haben.
2. Die Methode zur Zustandsbeurteilung ist nicht die richtige.

Da von SIDIS W hauptsächlich Parameter des Antriebs aufgezeichnet werden, ist es naheliegend, dass nicht alle relevanten Größen erfasst werden. Wie bereits erwähnt, wirken rund 75 Einflussparameter auf den Zustand einer Weiche [JZ02]. Insbesondere wenn die fehlenden Größen nicht zur Stellstromüberwachung gehören, passt auch die Methode nicht. Spätestens dann muss ein Zustandsmodell auch nicht lineare Zusammenhänge berücksichtigen. Bevor die Methode diskutiert werden kann, muss demzufolge untersucht werden, ob die richtigen Größen erfasst sind. In diesem Abschnitt wird daher die Frage beantwortet, welche zustandsbeeinflussenden Größen zusätzlich zur Datenbasis gehören sollten. Dazu werden zunächst einige Attribute näher beschrieben, die SIDIS W aufzeichnet (Abschnitt 5.2.1), damit die analysierten Zusammenhänge mit weiteren Größen in den Abschnitten 5.2.2 und 5.2.3 nachvollziehbar sind.

5.2.1 Detaillierte Beschreibung wesentlicher von SIDIS W gemessener Parameter

SIDIS W zeichnet insgesamt 56 Parameter auf. Die wichtigsten sind hier kurz erläutert. Hauptsächlich sind die Leistungsmittelwerte während der verschiedenen Umstellphasen (Pleer, Plauf1, Plauf2, Pversch1), der Spitzenwert am Beginn des Verschleißens (Pversch1PeakMax) und beim Abschalten der durchrutschenden Kupplung (Prutschmax) von Bedeutung. Abbildung 5-5 zeigt eine Umlaufkurve und die jeweiligen gemessenen Attribute, die in mittelbarem Zusammenhang mit dem Zustand der Weiche stehen. Darin ist auch die Umlaufzeit (Tumlauf) dargestellt.

Außerdem speichert SIDIS W weitere Werte, die Auskunft über den Zustand geben können. Dazu zählen elektrische Kennwerte wie die Stromasymmetrie (lunsym), die Motorspannung (Umotor) und die Beziehung zwischen Widerstand und Stromstärke (Rzul). Weitere Attribute sind die mechanischen Vorspannungen der beiden Zungen (Pzvorspg1, Pzvorspg2). Die Tabelle 5-5 listet die erwähnten Attribute samt ihrer Bedeutung auf, während eine komplette Aufstellung aller Attribute in der Tabelle B-1 im Anhang B.2 zu finden ist.

5.2.2 Klimabedingte Einflussfaktoren

Einige der Leistungswerte, die während des Weichenumlaufs aufgezeichnet werden, zeigen Auffälligkeiten im zeitlichen Verlauf, die je nach Weiche schwächer oder stärker ausgeprägt sind. Zum einen sind saisonale Schwankungen zu beobachten, zum anderen tägliche Schwankungen.

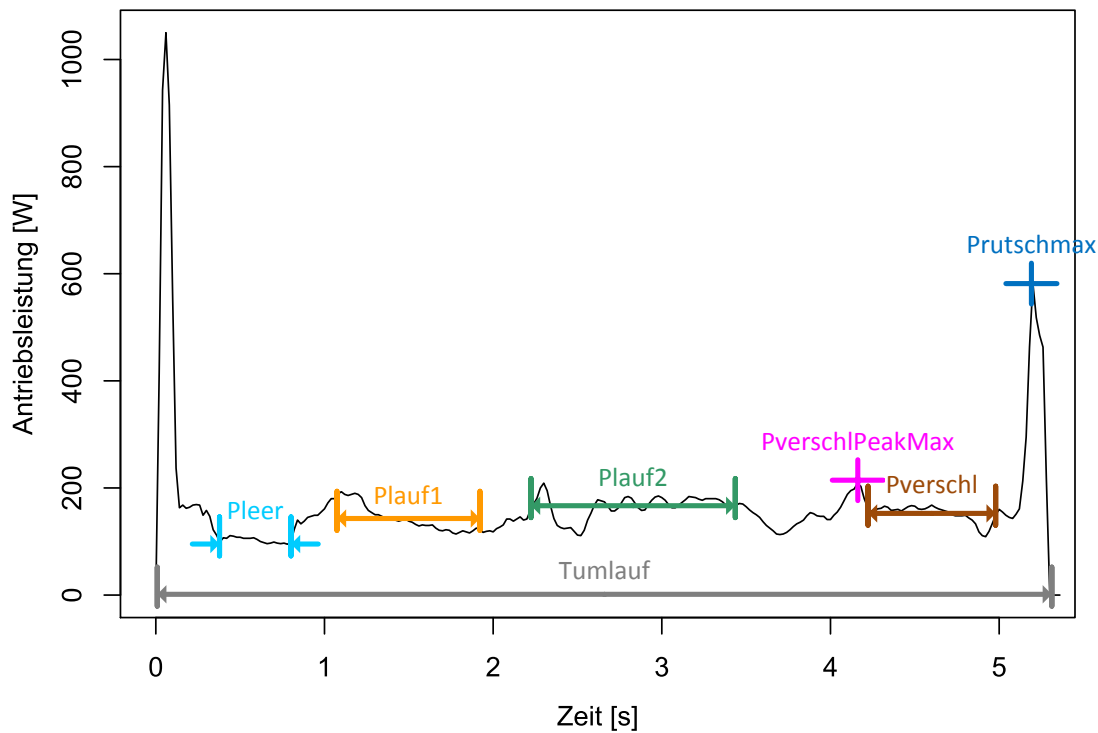


Abbildung 5-5: Beispiel eines von SIDIS W gemessenen Weichenumlaufs (siehe auch Abbildung 2-13); gekennzeichnet sind die Abschnitte der Mittelwertbildung für das jeweilige Attribut bzw. den Absolutwert der Umstelldauer

Attribut	Bedeutung
Rzul	Beziehung von Widerstand zu Stromstärke
Pleer	Leistungsdurchschnitt im Antriebsleerlauf
Plauf1	Leistungsdurchschnitt bei Zungenbewegung im Lauf der ablegenden Zunge
Plauf2	Leistungsdurchschnitt bei Zungenbewegung im Lauf beider Zungen
Pversch	Leistungsdurchschnitt während des Verschlussvorgangs
Tumlauf	Zeit des Weichenumlaufs
lunsym	Stromasymmetrie, der drei Phasen des Motorstroms
Pzvorspg1	Leistungsbeitrag durch mechanische Zungenvorspannung
Pzvorspg2	Leistungsbeitrag durch mechanische Zungenvorspannung
Prutschmax	Leistungsspitze beim Durchrutschen der Kupplung im Abschaltpunkt
Umotor	Motorspannung
PverschPeakMax	Leistungsspitze am Beginn der Verschlussphase

Tabelle 5-5: Ausgewählte Attribute, die in SIDIS W aufgezeichnet werden und deren Bedeutung

Die Abbildung 5-6 zeigt dies beispielhaft für die Weiche 7. Hier ist der Wert bei durchrutschen der Kupplung (Prutschmax) im Sommer deutlich niedriger als im Winter. Der Antriebsleerlauf (Pleer) und die abliegende Zungenbewegung (Plauf1) sind hingegen im Sommer höher als im Winter. Für die Werte der Bewegung beider Zungen (Plauf2), der Spitze am Verschlussbeginn (PverschPeakMax) und des Verschlusses (Pversch) sind die Effekte weniger ausgeprägt. Angesichts der starken Volatilität der Messwerte, ist auch die hohe Fehlalarmrate von SIDIS W nicht verwunderlich. Und noch ein Effekt ist zu beobachten. Nach der Störung im August sind die Werte für die Zungen- und Verschlussbewegungen deutlich abweichend. Die Ursache liegt in der Instandhaltungsmaßnahme bei der Entstörung, durch die die Weiche wieder leichtgängiger und der Verschluss gerichtet wurde.

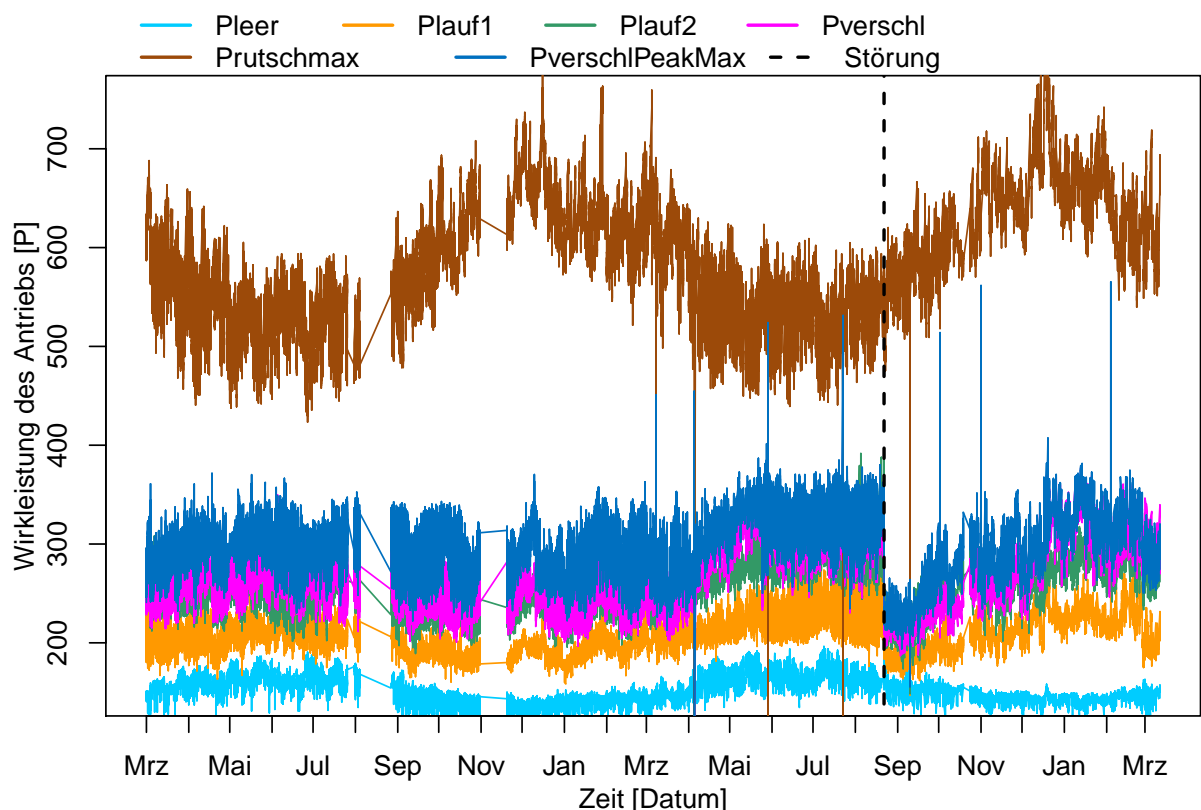


Abbildung 5-6: Verlauf einiger Leistungswerte der Weiche 7 während des Aufzeichnungszeitraums, mit auffälligen saisonalen und täglichen Schwankungen

5.2.2.1 In der Literatur benannte Klimaeinflüsse und akquirierte Daten

Das Klima scheint einen Einfluss auf den Zustand oder zumindest auf die Messungen zu haben. Diese Vermutung wurde auch in zahlreichen Veröffentlichungen zum Thema geäußert, allerdings ohne dies zu analysieren (vgl. [RB03], [Zwa09, S. 29], [Gut10, S. 108 f.] und [GMRT10, S. 40]). Einen konkreten Hinweis gibt nur das Handbuch von SIDIS W [Bol07, S. 104]. Darin wird auf Feuchtigkeit hingewiesen, die im Kabelkanal bzw. im Weichenantrieb für einen erhöhten Widerstand verantwortlich ist und dadurch die Stromasymmetrie (lunsym) beeinflusst. Im Rahmen der Forschung zur vorliegenden Arbeit wurden die klimatischen Zusammenhänge erstmals detailliert, datentechnisch untersucht und teilweise bereits veröffentlicht (vgl. [BD10, Böh12c]). Die hier folgenden Betrachtungen gehen allerdings darüber hinaus.

Als erstes mussten Klimadaten für den Zeitraum der SIDIS W Aufzeichnungen akquiriert wer-

den. Da solche Daten nicht vom Diagnosesystem und auch nicht von der DB Netz AG erhoben werden, wurde auf Daten öffentlich zugänglicher Wetterstationen zurückgegriffen. Die Nutzung dieser Daten ist nicht nur kostenlos, sondern erfüllt durch den Verzicht auf zusätzliche Sensoren auch die Anforderung, so wenig Sensoren wie möglich einzusetzen (siehe Nummer 5 in Tabelle 4-1). Die Daten stellen zwar eine externe Datenquelle dar, die aber im Sinne des KDD-Prozesses lediglich eine von vielen, zu integrierenden Quellen ist (vgl. Abbildung 3-1). Die Web-Site wetter.com stellt die Aufzeichnungen verschiedener Stationen zur Verfügung [wet09]. Die nächstgelegene Station ist nur fünf bis zehn km von den Weichen entfernt. Im halbstündlichen Rhythmus werden dort die Größen gemessen, die in der Tabelle 5-6 aufgelistet sind. Um die Daten der Webseiten nutzen zu können, wurde eigens ein Programm (sogenannter Webcrawler) geschrieben, das die Webseiten durchsucht, die relevanten Daten extrahiert und so speichert, dass sie später weiterverarbeitet werden können.

Attribut	Bedeutung
WErfDatum	Datum und Uhrzeit der Wetterdatenerfassung
Feuch in %	Luftfeuchtigkeit
Temperatur in °C	Lufttemperatur
Luftd in hPa	Luftdruck
Windr in °	Windrichtung
Nieder in mm	Niederschlag
Winds in kn	Windgeschwindigkeit
Sicht in m	Sichtweite

Tabelle 5-6: Attribute der Wetterstation und ihre Bedeutung

5.2.2.2 Analyse des Zusammenhangs mit Klimadaten

Die extrahierten Klimadaten wurden anhand der Aufzeichnungszeitpunkte mit den Messdaten aus SIDIS W verknüpft und auf ihre lineare Korrelation (Pearson Korrelation) hin analysiert. Da die Weichen unterschiedlichen Typs sind und verschieden kalibriert werden können, verhalten sie sich auch unterschiedlich. Deshalb wurden die Korrelationstests für jede Weiche und jede Umstellrichtung durchgeführt. Wegen des unverhältnismäßig hohen Platzbedarfs, wird hier von allen 58 Analysen das arithmetische Mittel der jeweiligen Korrelationskoeffizienten präsentiert, um auffällige statistische Zusammenhänge darzustellen. Das Ergebnis zeigt Abbildung 5-7. Sich kreuzende Zeilen und Spalten zeigen den durchschnittlichen Koeffizienten des jeweiligen Attributpaares. Die linke untere Hälfte enthält dabei die Werte, während rechts oben allein die graphische Repräsentation dargestellt ist.

Was im zeitlichen Verlauf der Messwerte bei Weiche 7 vermutet wurde (siehe Abbildung 5-6), bestätigt sich hier.

- Vor allem Pleer (0.6) und Prutschmax (-0.77) korrelieren mit der Temperatur. Fällt diese, so sinkt die Leerlaufleistung bzw. steigt die Wirkleistung des Motors im Abschaltpunkt und umgekehrt. Der statistische Zusammenhang lässt sich allgemein mit temperaturabhängigen Materialeigenschaften erklären. Der Motor kann bspw. mehr Spiel besitzen und leichtgängiger sein, während die Reibung der Kupplung bei Kälte höher ist. Ohne dass die Korrelation es zeigt, können zusätzlich die Zungen und der Verschluss beeinflusst werden, weil sie sich ausdehnen oder zusammenziehen. Außerhalb einer gewissen Toleranz kann

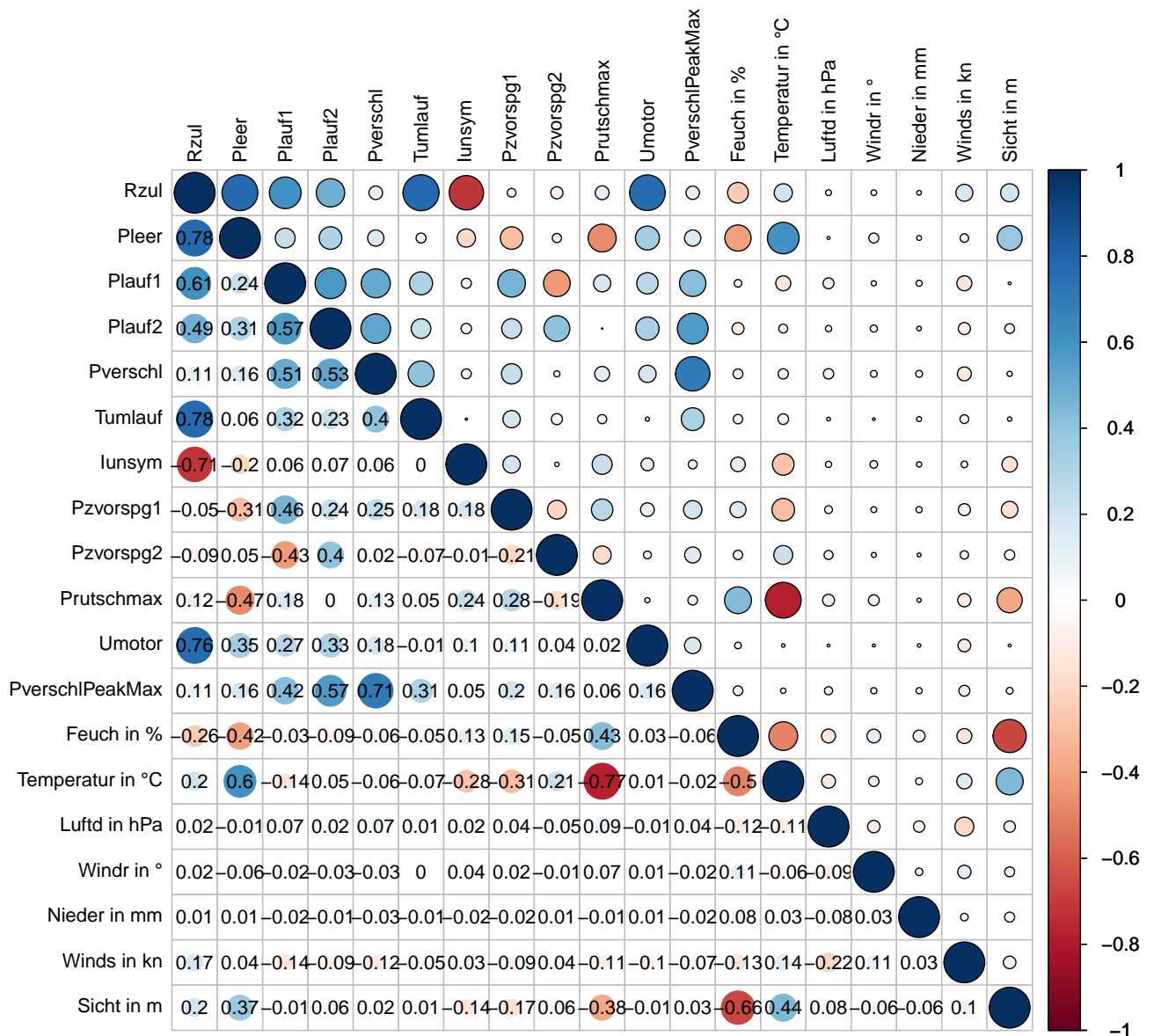


Abbildung 5-7: Arithmetisches Mittel aller linearen Korrelationskoeffizienten je Weiche und Umstellrichtung

dies zu Verkantungen im Verschluss und letztlich zur Blockierung der Weiche führen (vgl. Zungenwanderung beim Verschluss im Abschnitt 2.1.2).

- Auch die Korrelation dieser beiden Attribute mit der Luftfeuchtigkeit (Pleer -0.42 und Prutschmax 0.43) ist auffällig, wenn auch nicht so ausgeprägt wie bei der Temperatur. Die Temperatur und die Luftfeuchtigkeit stehen jedoch in einem physikalischen Zusammenhang, was sich in dem Korrelationskoeffizienten von -0.5 zwischen beiden spiegelt. Es kann also nicht genau gesagt werden, ob der Zusammenhang mit SIDIS W nur statistisch existiert. Der in der Literatur benannte Zusammenhang zwischen Luftfeuchtigkeit und Stromasymmetrie (lunsym) ist hier nicht belegbar.
- Als drittes besitzt noch die Sichtweite nennenswerte Korrelationswerte. Ein kausaler Zusammenhang zum Weichenzustand besteht jedoch nicht. Da die Sicht hauptsächlich das Resultat von Wassertröpfchen und Staub in der Atmosphäre ist, lässt sich nicht erklären,

wie diese den Weichenzustand direkt beeinflussen könnte. Temperatur und Luftfeuchtigkeit spielen für die Sichtweite eine entscheidende Rolle, wie deren Korrelation zeigt. Das führt dann logischerweise zu der Korrelation mit SIDIS W Attributen. Entsprechend handelt es sich um eine Koinzidenz.

Von den Attributen der Klimadaten sind also nur die Temperatur und die Luftfeuchtigkeit relevant. Wie oben erwähnt, kann das Verhalten der Weichen unterschiedlich sein. Weil hier der Durchschnitt der Korrelationskoeffizienten verwendet wurde, ist noch wichtig zu sehen, wie dieser Mittelwert streut. Schließlich ist bspw. der Mittelwert einer Sinusfunktion 0, obwohl dieser Funktionswert nur selten zutrifft. Aus dem Boxplot in Abbildung 5-8 wird die Verteilung der Korrelationswerte für die Temperatur (a) und die Luftfeuchtigkeit (b) ersichtlich. Der Boxplot wird hier wegen der besseren Übersichtlichkeit verwendet, während die Werte der Standardabweichung in Tabelle B-4 im Anhang B.3 zu finden sind. Die Box, also das obere und untere Quantil von Pleer und Prutschmax liegen deutlich und eng auf positiver bzw. negativer Seite. Selbst die Ausreißer sind begrenzt. Die Korrelation ist damit eindeutig. Rzul, lunsym und Umotor bleiben in engem Rahmen unkorreliert. Andere Attribute wie die Bewegung der abliegenden Zunge (Plauf1), die Bewegung beider Zungen (Plauf2), das Verschließen (Pversch1) und die Umlaufzeit (Tumlauf) streuen stark. Die Korrelation ist aber überwiegend gering und auf positiver wie negativer Seite annähernd gleich. Es ist zu bedenken, dass Plauf1 und Pversch1 je Umstellrichtung wechseln. Was in die eine Richtung leicht zu bewegen ist, kann in der anderen schwergängig sein. Das Wandern der Zungen oder ähnliches kann dann je Richtung gegensätzliche Effekte haben, was die lineare Korrelation nicht erfassen würde. Demzufolge kann eine Temperaturabhängigkeit bei der Zungenbewegung und dem Verschluss weder bestätigt noch widerlegt werden. Als Schlussfolgerung aus der Korrelationsanalyse geht hervor, dass nicht das Klima allgemein, sondern konkret nur die Temperatur sowie die Luftfeuchtigkeit bei der Zustandsprognose zu berücksichtigen sind. Beide Parameter werden meteorologisch vorhergesagt, wodurch diese Werte auch in die Störungsprognose integriert werden können.

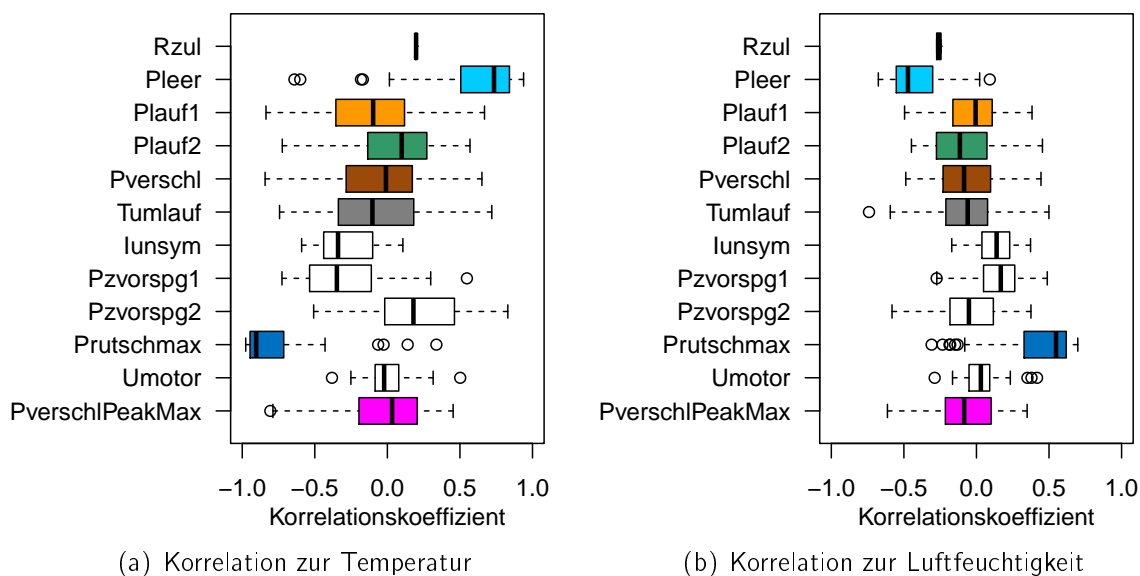


Abbildung 5-8: Boxplot der Korrelationskoeffizienten von SIDIS W Attributen zur Temperatur und zur Luftfeuchtigkeit

Im Zusammenhang mit der ausgeprägten Temperaturkorrelation sei noch ein Experiment er-

wähnt. Wenn die starken Schwankungen der Messwerte aus Temperaturänderungen resultieren, könnte dieser Rauscheffekt dann nicht einfach herausgefiltert werden? Die Korrelation mit der Temperatur kann dazu benutzt werden, um deren Einfluss aus den Messgrößen herauszurechnen. Dadurch schwanken diese weniger und verletzen seltener die Alarmgrenzen, wodurch letztlich auch Fehlalarme reduziert werden. Bei elf Weichen konnten so die gelben Alarme auf 14134 von 16997 (83.1 %) reduziert werden, die roten sogar auf 386 von 4724 (8.1 %). Dies wurde in den Veröffentlichungen von Böhm [Böh12a, Böh12b] ausführlich erläutert. Auf eine Wiederholung wird hier verzichtet, da die zu Abschnitt 5.1.4 analoge Performance-Analyse mittels ROC-Graph ergeben hat, dass sich die Genauigkeit in Bezug auf eine Zustandsbeurteilung nicht im ausreichenden Maße verbessert. Es werden nicht deutlich mehr Störungen richtig erkannt und bei den vielen Messungen sinkt auch die Fehlalarmrate nur geringfügig (vgl. Abbildung B-2 in Anhang B.3). Außerdem ist das einfache, temperaturabhängige Glätten von Messwerten nicht sinnvoll. Denn bei Temperaturänderungen dehnen bzw. verkürzen sich die Zungen, wodurch der Verschluss verkanten kann und die Weiche blockiert. Beim Glätten der Messwerte würde diese Störungsursache verdeckt. Vielmehr wird daher das Fazit des Experiments befolgt, das dazu rät, Temperaturdaten aus externen Datenquellen in die Zustandsprognose zu integrieren. [Böh12a]

5.2.3 Bauartbedingte Einflussfaktoren

Es wurde bereits mehrfach erwähnt, dass Weichen sich unterschiedlich verhalten. So zeigen sie unterschiedlich ausgeprägte Korrelationswerte mit der Temperatur und der Luftfeuchtigkeit. Auch die Antriebsmesswerte haben unterschiedliche Niveaus oder schwanken in verschiedenen Intervallen. Dies ist bei einer Zustandsprognose zu berücksichtigen. Denn was bspw. bei der einen Weiche als kritischer Leistungswert während der Zungenbewegung anzusehen ist, kann bei einer anderen Weiche völlig unproblematisch sein. Diese Effekte werden im Folgenden untersucht.

5.2.3.1 In der Literatur benannte Bauarteinflüsse und akquirierte Daten

Nach verschiedenen Literaturquellen zu urteilen, liegen die Gründe für dieses unterschiedliche Verhalten zum Teil in den Weichentypen und ihren vielfältigen Konstruktionselementen. Hauptsächlich ist hier die Dissertation von Zwanenburg [Zwa09] zu nennen. Seine Arbeit beschäftigt sich intensiv mit dem Abnutzungsverlauf von Weichen. Obwohl Zwanenburg selbst festhält, dass der Antrieb Hauptstörungsursache ist [Zwa09, S. 102], schließt er diesen leider vollständig aus. Ungeachtet dessen hat er herausgefunden, dass die Alterung der beiden untersuchten Weichentypen einfache Weiche (EW) und doppelte Kreuzungsweiche (DKW) differenziert verläuft [Zwa09, S. 93]. Dieses Ergebnis wird mit der Veröffentlichung von Thiel bestätigt [Thi10], nach der die aufwendiger konstruierten Kreuzungs-, Innenbogen- (IBW) und Außenbogenweichen (ABW) schneller verschleifen bzw. mehr Instandhaltung benötigen. Die Schwellenart und die Unterbauqualität beeinflussen ebenfalls die Alterung von Weichen [Zwa09, S. 57]. Und auch der Radius und der Abzweigwinkel spielen eine Rolle [Zwa09, S. 79 und 82]. Weichen mit kleinem Radius sind höheren Belastungen durch die seitlichen Kräfte bei der Zugüberfahrt ausgesetzt (vgl. Abschnitt 2.1.2). Gleiches gilt auch für die Geschwindigkeit, mit der die Weichen befahren werden können, die wiederum vom Radius bzw. vom Winkel abhängt. Je höher die Geschwindigkeit ist, desto geringer ist die Lebensdauer einer Weiche [Zwa09, S. 87 f.].

Neben dem Weichentyp, der Schwellenart, der Unterbauqualität und dem Radius bzw. dem Ab-

zweigwinkel beeinflussen auch die Antriebsart und die Verschlussart das Zustandsverhalten. Die Verschlussart ist wegen der unterschiedlichen Toleranz gegenüber der Zungenwanderung von Bedeutung, nach der der Verschluss mehr oder weniger anfällig für Störungen durch Verklemmung ist. Das Phänomen wurde bereits im Abschnitt 2.1.2 bei der Beschreibung der Verschlussarten und im Abschnitt 5.2.2 bei den klimabedingten Einflussfaktoren erläutert. Unterschiedliche Antriebsarten haben unterschiedliche Störungsschwerpunkte und verschleßen auf verschiedene Weise. Im Vergleich zu elektrohydraulischen Antrieben können z.B. bei elektromechanischen Bauformen Getriebe und Kupplung verschleßen [Uni08, S. 16], [SL08, S. 11]. Wenn Weichen aufgrund ihrer Länge mehrere Antriebe und Verschlüsse haben, kommt es zudem vor, dass eine Funktionsschwäche eines Antriebs durch die anderen kompensiert wird [Böh13c]. Dadurch wären natürlich auch die Leistungsmesswerte sehr unterschiedlich. Ein weiterer Aspekt der Bauart, welcher zu beachten ist, ist die Verwendung von Zungenrollern [Gut10, S. 112]. Zungenroller verringern den Reibungswiderstand bei der Zungenbewegung (vgl. Abschnitt 2.1.2), wodurch auch die Leistungswerte des Antriebs auf einem anderen Niveau liegen [GB11, S. 579].

All diese Bauarteigenschaften von Weichen können das unterschiedliche Verhalten erklären, welches sich in den SIDIS W Messdaten findet. Es ist allerdings nicht klar, welche Bedeutung den verschiedenen Konstruktionselementen beizumessen ist. Ebenso wie bei den klimabedingten Einflüssen, existiert keine systematische Analyse darüber, welche Parameter mit welchen Messgrößen in Zusammenhang stehen. Aber gerade für die Allgemeingültigkeit der Zustandsprognose ist das wichtig, weil das zu entwickelnde Prognosemodell ja auf verschiedene Weichen anwendbar sein soll. Daher wird der Zusammenhang hier erstmalig untersucht.

Als Quelle für die Bauarteigenschaften der 29 Weichen wurden ebenfalls von der DB Netz AG bereitgestellte Daten aus SAP genutzt, die sogenannten Weichenstammdaten. Die Tabelle 5-7 enthält die jeweiligen Charakteristiken der Weichen, in denen sie sich unterscheiden. Zudem besitzen alle Weichen das Schienenprofil S54, den Antriebstyp S700K und den gleichen Endlagenprüfer (ELP). Insgesamt zeigt die Tabelle nicht nur, dass sehr unterschiedliche Weichen untersucht werden, sondern auch, dass sich darunter die sehr häufig im Netz der DB AG verbauten Einzelweichen mit Klinkenverschluss und S700K-Antrieb befinden. Dementsprechend sind zwei weitere Anforderungen an das zu entwickelnde Prognoseverfahren erfüllt (Weichen mit verschiedenen Eigenschaften und häufig verbaute Weichentypen verwenden, siehe Anforderung Nummer 7 und 9 der Tabelle 4-1).

5.2.3.2 Verfahren zur Analyse des Zusammenhangs mit Bauarteigenschaften

Die Analyse des Zusammenhangs zwischen SIDIS W Messgrößen und Bauarteigenschaften ist aufwendiger als bei den klimabedingten Einflüssen. Denn im Gegensatz zu diesen kann keine Korrelationsanalyse genutzt werden, weil die Weichencharakteristik über den Messzeitraum konstant bleibt. Andere statistische Analyseverfahren, wie etwa Hotellings T^2 -Test, Wilcoxon-Mann-Whitney-Test, Varianzanalyse (ANOVA) oder Diskriminanzanalyse können ebenfalls nicht verwendet werden. Die mit diesen Verfahren verbundenen Voraussetzungen sind nicht erfüllt. Denn die SIDIS W Attribute sind nicht unabhängig voneinander. Die Messwerte der Attribute sind teilweise nicht normalverteilt, und die Varianzen der Attribute sind verschieden (vgl. Abbildung B-3 im Anhang B.4). Aus diesem Grund wird ein eigenes heuristisches Verfahren verwendet. Die Heuristik beruht auf der aus den oben genannten Literaturquellen abgeleiteten Annahme,

Weiche	Weichentyp	Abzweigrichtung	Radius in m	Winkel 1:m	Schwellenart	Verschlussart	Anzahl Zungenrollpaare
2	IBW	L	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	6
1	IBW	L	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	6
21	IBW	L	500	12	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	7
4	EW	L	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	7
3	EW	L	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	7
10	EW	L	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	7
16	EW	R	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	7
7	EW	R	1200	18.5	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	5
19	EW	R	500	14	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
18	EW	R	500	14	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
25	EW	L	500	14	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
5	EW	L	760	14	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
29	EW	R	500	14	Holz	Klinkenmittelverschluss 2 Klinkenspitzenverschluss	5
28	EW	R	500	14	Holz	Klinkenmittelverschluss 2 Klinkenspitzenverschluss	5
23	EW	L	500	14	Holz	Klinkenmittelverschluss 2 Klinkenspitzenverschluss	5
22	IBW	L	500	12	Beton	Klinkenmittelverschluss 2 Klinkenspitzenverschluss	5
14	EW	L	1200	18.5	Beton	Gabelmittelverschluss 1 Klinkenmittelverschluss 1 Klinkenspitzenverschluss	0
11	EW	R	1200	18.5	Beton	Gabelmittelverschluss 1 Klinkenmittelverschluss 1 Klinkenspitzenverschluss	0
13	EW	L	1200	18.5	Beton	Gabelmittelverschluss 1	0
9	EW	L	1200	18.5	Beton	Gabelmittelverschluss 1	0
12	EW	R	1200	18.5	Beton	Gabelmittelverschluss 1	0
8	EW	R	1200	18.5	Beton	Gabelmittelverschluss 1	0
15	EW	R	1200	18.5	Beton	Gabelmittelverschluss 1	0
6	ABW	R	190	9	Holz	Klinkenspitzenverschluss	2
27	EW	L	500	14	Holz	Klinkenmittelverschluss 2 Klinkenspitzenverschluss	5
26	EW	L	500	14	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
24	EW	L	500	12	Beton	gleit.Verschl.-Stück SpV gleit.Verschl.-Stück MV	0
20	EW	R	500	12	Holz	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4
17	EW	L	500	12	Beton	Klinkenmittelverschluss 1 Klinkenspitzenverschluss	4

Tabelle 5-7: Bauarteigenschaften der 29 Weichen (Die ersten 24 Zeilen sind nach Ähnlichkeit sortiert. Außerdem besitzen alle Weichen das Schienenprofil S54, den Antriebstyp S700K und den gleichen Endlagenprüfer.)

dass Weichen mit ähnlichen Eigenschaften auch ein ähnliches Verhalten in den Messdaten zeigen. Es gilt nun sowohl die relevanten Eigenschaften als auch die relevanten Messattribute zu identifizieren, die diese Annahme bestätigen oder die gesamte Annahme zu widerlegen. Dazu werden die folgenden Schritte durchgeführt:

1. **Auswählen einer passenden Datenuntermenge:** Um Einflüsse des Wetters o. ä. auszuschließen, wird zuerst eine Untermenge der Messdaten und Weichen ausgewählt, die im gleichen und möglichst großen Zeitraum gemessen wurden. Dafür ist der Zeitraum von Mitte September 2008 bis Mitte Februar 2009 am besten geeignet. Allerdings müssen die Weichen mit den Nummern 17, 20, 24, 26, 27 mangels Messdaten ausgeschlossen werden (vgl. Abbildung 5-1). Die verbleibenden 24 Weichen enthalten in den fünf Monaten aber immer noch 78336 Messungen ($\mathcal{S}_{Sep08Feb09} \subset \mathcal{S}$, $|\mathcal{S}_{Sep08Feb09}| = 78336$). Die Auswahl der Messattribute umfasst die Zustandsparameter aus Tabelle 5-5 (Pleer, Plauf1, Plauf2, Pverschl, Tumlauf, lunsym, Pzvorspg1, Pzvorspg2, Prutschmax, Umotor, PverschlPeakMax). Das verbleibende Attribut Rzul ist konstant für nahezu alle Messungen je Weiche, wodurch es zwar für die Identifikation einer Weiche geeignet ist – fast so gut wie die Weichenbezeichnung selbst – nicht aber für die Identifikation eines Verhaltenszusammenhangs.
2. **Gruppieren von Weichen mit ähnlichen Bauarteigenschaften mittels Clustering:** Clustering Verfahren finden in der KDD Anwendung, wenn in Daten unbekannte Gruppen identifiziert werden sollen [FPSSU96]. Hauptsächlich werden Distanzmaße zwischen den Objekten gebildet, in diesem Fall zwischen den Weichen. Anschließend werden jene Weichen in einer Gruppe zusammengefasst, deren Distanzmaß im Verhältnis zu den anderen Weichen klein ist. Für die Distanzberechnung und für die Bildung von Gruppen gibt es eine ganze Reihe von Algorithmen, mit verschiedenen Vor- und Nachteilen bzw. für spezielle Anwendungsbereiche. Da diese jedoch nicht im Fokus der vorliegenden Arbeit liegen, sei zur Vertiefung auf die umfassenden Grundlagenwerke von Petersohn [Pet05], Theodoridis und Koutroumbas [TK09] oder Linoff und Berry [LB11] verwiesen. Die Attribute der Bauart besitzen ordinalskalierte Werte (Radius, Winkel, Zungenroller) und nominalskalierte Werte (Typ, Richtung, Schwelle, Verschluss). Darum wird hier zum Clustering das Distanzmaß nach Gower [Gow71] verwendet, welches eines der meist angewendeten ist. Für die Bildung der Gruppen wird ein hierarchisches, agglomeratives Clustering genutzt. Beginnend bei je einem Cluster für jede einzelne Weiche werden dabei stufenweise Cluster zu einem neuen zusammengefasst, die sich jeweils am ähnlichsten sind, bis letztlich alle Weichen in einem Cluster vereint sind. Dadurch kann leicht die Anzahl der Gruppen geändert werden, ohne dass sich die Struktur und Beziehung der Weichen untereinander ändert. Das Ergebnis ist in Form eines Dendrogramms (Baum der Zerlegung bzw. Zusammenfassung) in Abbildung 5-9 dargestellt. Entlang des Dendrogramms werden zwei bis maximal 16 Gruppen gebildet. Innerhalb der 16 Cluster sind dann die Bauarteigenschaften gleich, so dass eine weitere Unterteilung nur noch die 24 Weichen ergibt (G^v , $v \in V = \{2, \dots, 16, 24\}$).
3. **Klassifikation der Bauartgruppenzugehörigkeit mittels Messdaten:** Wenn überhaupt ein Zusammenhang zwischen der Bauart und dem Weichenverhalten besteht, dann muss es Bauartcluster geben, die sich mit Hilfe der Messdaten bestimmter SIDIS W Attribute identifizieren lassen. Bspw. könnten Gruppen mit unterschiedlicher Zungenrollpaaranzahl, wegen des unterschiedlichen Reibungswiderstands, auch überwiegend unterschiedliche Werte bei der Bewegung beider Zungen (Plauf2) aufweisen. Auch wenn sich die Wertebereiche

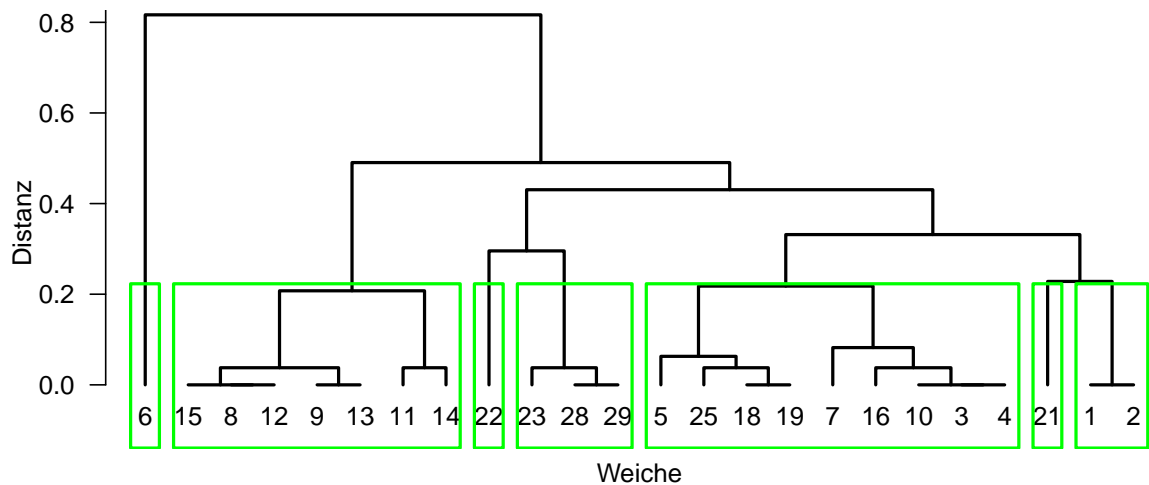


Abbildung 5-9: Dendrogramm der Weichen und ihrer Bauarteigenschaften gemäß des Distanzmaßes nach Gower für das hierarchische, agglomerative Clustering bspw. ergäben sich bei sieben Clustern die hier grün umrandeten Gruppen)

der Gruppen überschneiden, sollte es möglich sein, anhand von Plauf2 zu bestimmen, welcher Gruppe die Messung angehört – sofern die Annahme korrekt ist. Allgemein ist zu beachten, dass die meisten Clustering Algorithmen immer terminieren, d.h. immer eine Gruppierung finden, ungeachtet dessen, ob diese Cluster tatsächlich eine signifikante Struktur wiedergeben [Mil85]. Als Signifikanztest wird daher der Vergleich des Clusterings gemäß Eigenschaftsdaten mit einer rein zufälligen Gruppierung herangezogen. Für die vorliegende Untersuchung bedeutet dies, dass eine Klassifikation der Bauartcluster deutlich genauer sein muss, als eine Klassifikation zufällig zusammengestellter Gruppen von Weichen. Nur dann kann die gefundene Gruppenstruktur als aussagekräftig gelten [HBV01]. Gleichzeitig sollte die Klassifikation der Bauartcluster genauer sein, als die Klassifikation der einzelnen Weichen (24 Cluster) (G^{24}). Im Einzelnen wird die Klassifikation in 16 Iterationen (mögliche Clusteranzahl $|V| = 16$) folgendermaßen durchgeführt:

- Es wird künstlich eine Anzahl (v) von Gruppen gebildet, denen allerdings die Weichen zufällig zugewiesen werden, um so die Signifikanz der identifizierten Gruppenstruktur bzw. der Klassifikation zu prüfen.
- Die Menge der ausgewählten Messdaten ($\mathcal{S}_{Sep08Feb09}$) wird in eine zufällige Reihenfolge gebracht, damit nicht etwa die geordneten Messdaten einen Einfluss auf die Klassifikation ausüben.
- Je gewählter Gruppenanzahl (v) wird für jede Gruppe eine gleich große Menge aus dem Datensatz ($\mathcal{S}_{Sep08Feb09}$) gezogen ($|G_h^v| = |G_i^v|$, $\forall h, i \in \{2, \dots, v\}$), damit die Klassifikation nicht von der Gruppengröße abhängt.
- Da nicht bekannt ist, welche der SIDIS W Attribute (siehe Tabelle 5-5) oder welche ihrer Kombinationen im Zusammenhang mit den Bauarteigenschaften stehen, müssen alle möglichen Kombinationen analysiert werden. D.h. die Klassifikation muss für jede Teilmenge der Potenzmenge der elf Attribute durchgeführt werden, ausschließlich der leeren Menge. Es ergeben sich 2047 ($2^{11} - 1$) Iterationen der folgenden Teilschritte:
 - Die Datenmenge wird zufällig in eine 70 Prozent umfassende Trainingsmenge und eine 30 Prozent umfassende Testmenge zerlegt.

- ii. Unter Verwendung eines Bayes'schen Netzes (siehe Abschnitt 3.2.2) werden Klassifikationsregeln anhand der Trainingsmenge erlernt. Anschließend werden die Regeln anhand der Testmenge evaluiert, die während des Trainings unbekannt geblieben ist. Dafür wird die Konfusionsmatrix der Klassen (hier die Clusterzugehörigkeit) mit dem Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient) (MCC) (siehe Abschnitt 3.3.2) ausgewertet. Der MCC ist an dieser Stelle die bevorzugte Metrik, weil er kleine Unterschiede in der Klassenzugehörigkeit anzeigen kann, für zwei und mehr Klassen anwendbar und zudem bei Rauschen innerhalb der Klassen geeignet ist (vgl. Tabelle 3-8).

Diese Klassifikation und ihre Bewertung werden parallel jeweils für die Bauartcluster und die zufällig zusammengestellten Cluster durchgeführt. Dadurch wird die Qualität anhand des MCC verglichen und die Signifikanz der Bauartklassifikation auf Basis bestimmter SIDIS W Attribute bewiesen oder widerlegt.

4. Der dritte Schritt wird ausreichend oft (hier 100 Mal) wiederholt, um auszuschließen, dass das Klassifikationsergebnis auf einer günstigen oder ungünstigen Datenauswahl, Datenreihenfolge o. ä. beruht. Letztlich werden das arithmetische Mittel und die Standardabweichung der MCC-Werte errechnet.

5.2.3.3 Ergebnisauswertung des Analyseverfahrens

Nachdem die eben beschriebenen Schritte ausgeführt wurden, werden die Ergebnisse wie folgt beurteilt. Die Standardabweichung des arithmetischen Mittels des MCC, welches aus den 100 Wiederholungen je Clusteranzahl und Attributkombination berechnet wurde, zeugt von insgesamt stabilen Klassifikationsergebnissen. Die Standardabweichung ist in keinem Fall größer als 0.086, und in 99 Prozent aller Mittelwertbildungen ist sie sogar kleiner als 0.035. Das eigentliche Ergebnis der Klassifikation ist in Abbildung 5-10 dargestellt. Es zeigt, dass die Klassifikationserfolge mit der Clusteranzahl variieren. Noch deutlicher vermittelt die Abbildung die Unterschiede zwischen den 2047 Attributkombinationen. Einige erreichen sehr gute MCC-Werte oberhalb von 0.8 (1.0 repräsentiert eine perfekte Klassifikation, -1.0 eine vollkommen falsche). Andere Werte liegen unterhalb von 0.2. Letzteres bedeutet, dass die Clusterzugehörigkeit eher gleich verteilt erfolgte, also in jeder Zelle der Konfusionsmatrix annähernd gleich viele Messungen zusammenfielen. In diesen Fällen war eine Klassifikation der Bauartclusterzugehörigkeit auf Basis der betreffenden Messattribute nicht erfolgreich. Insgesamt beweist die Abbildung, dass einige Attribute stark im Zusammenhang mit den Bauarteigenschaften stehen.

Werden für jede Clusteranzahl (v) jeweils die schlechtesten 50 und die besten 50 MCC-Mittelwerte betrachtet, so wie in der Tabelle 5-8, lässt sich die Interpretation detaillieren. Bei den schlechtesten Klassifikationsergebnissen waren hauptsächlich nur ein bis drei Attribute beteiligt. Außerdem sind darin besonders oft die Umlaufzeit (Tumlauf) und die Zungenvorspannungen (Pzvorspg1, Pzvorspg2) enthalten. Selten wurden die Leerlaufleistung (Pleer), die Stromasymmetrie (lunsym) sowie das Verschlussmaximum (PverschPeakMax) verwendet, wenn das Klassifikationsergebnis schlecht ausfiel. Andersherum wurden gute Klassifikationsergebnisse überwiegend mit sieben bis zehn Attributen erreicht. Sehr häufig waren die Attribute Pleer, die Verschlussleistung (Pverschl), lunsym und PverschPeakMax involviert. Eher Selten waren es Tumlauf und Pzvorspg1. Diesen Ergebnissen folgend, sind es hauptsächlich Attribute des Motors und des

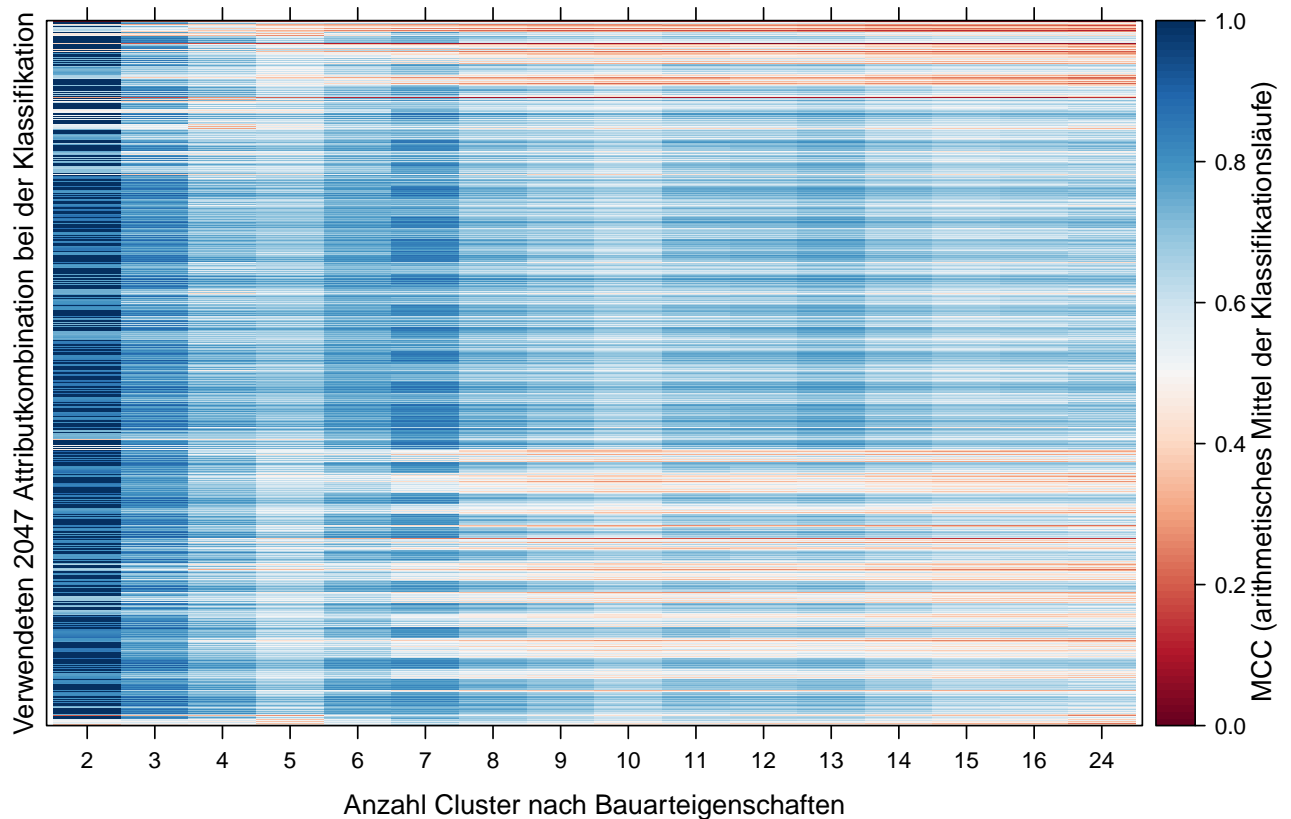


Abbildung 5-10: MCC der Klassifikation der Bauartclusterzugehörigkeit nach SIDIS W Messdaten

Anzahl enthaltener Attribute	Häufigkeit in den 50 schlechtesten besten MCC je Clusteranzahl	
1	141	-
2	387	-
3	231	1
4	37	5
5	4	16
6	-	35
7	-	78
8	-	222
9	-	268
10	-	147
11	-	28

Attribut	Häufigkeit in den 50 schlechtesten besten MCC je Clusteranzahl	
Pleer	58	693
Plauf1	111	514
Plauf2	104	500
Pversch1	188	797
Tumlauf	303	432
lunsym	70	744
Pzvorspg1	226	432
Pzvorspg2	288	638
Prutschmax	169	616
Umotor	200	668
PverschPeakMax	59	791

Tabelle 5-8: Häufigkeit von Kombinationen mit einer bestimmten Attributanzahl (links) und Häufigkeit eines bestimmten Attributs (rechts) in den 50 schlechtesten und besten Klassifikationsergebnissen (MCC-Mittelwert) je Clusteranzahl

Verschlusses (unterstützt von Attributen der Zungenbewegung), mit denen die Bauartcluster klassifiziert werden. Es sind in erster Linie diese Attribute, die den Zusammenhang zwischen Bauarteigenschaften und Weichenverhalten herstellen. Welche Bauarteigenschaften nun bei der Entwicklung des Prognosemodells zu berücksichtigen sind, legt die folgende Auswertung offen.

Neben den Unterschieden im Klassifikationsergebnis, die auf die Attributkombinationen zurückzuführen sind, unterscheidet sich auch das Klassifikationsergebnis bei variierender Clusteranzahl. Zur besseren Übersicht sind die MCC-Werte je Clusteranzahl in Abbildung 5-11 dargestellt. Bei zwei Clustern können die zugehörigen Weichen mitunter eindeutig ($MCC = 1.0$) durch SIDIS W Messdaten klassifiziert werden. Dies ist jedoch nicht verwunderlich, da sich in einem Cluster nur die Weiche 6 befindet. Alle restlichen Weichen sind im zweiten Cluster zusammengefasst (vgl. Abbildung 5-9). Als einzige Außenbogenweiche, die noch dazu die einzige mit nur einem Klinkenspitzenverschluss ist und den kleinsten Radius und Winkel besitzt, hat Weiche 6 kaum gemeinsame Bauarteigenschaften mit anderen Weichen (vgl. Tabelle 5-7). Eine Weiche vom Rest zu unterscheiden, ist allerdings kaum hilfreich, wenn es darum geht, ein allgemeingültiges Prognosemodell zu entwickeln. Außerdem lässt die Lücke zwischen den MCC-Werten von 1.0 und rund 0.9 den Schluss zu, dass es Attributwerte gibt, die insbesondere die Weiche 6 identifizieren, aber bei weiteren Gruppierungen nicht verwendbar sind. Es besteht also die Gefahr, dass dabei nur eine Ausnahme vorliegt. Es ist daher wünschenswert, auch die Gruppe der verbleibenden Weichen weiter aufzuschlüsseln.

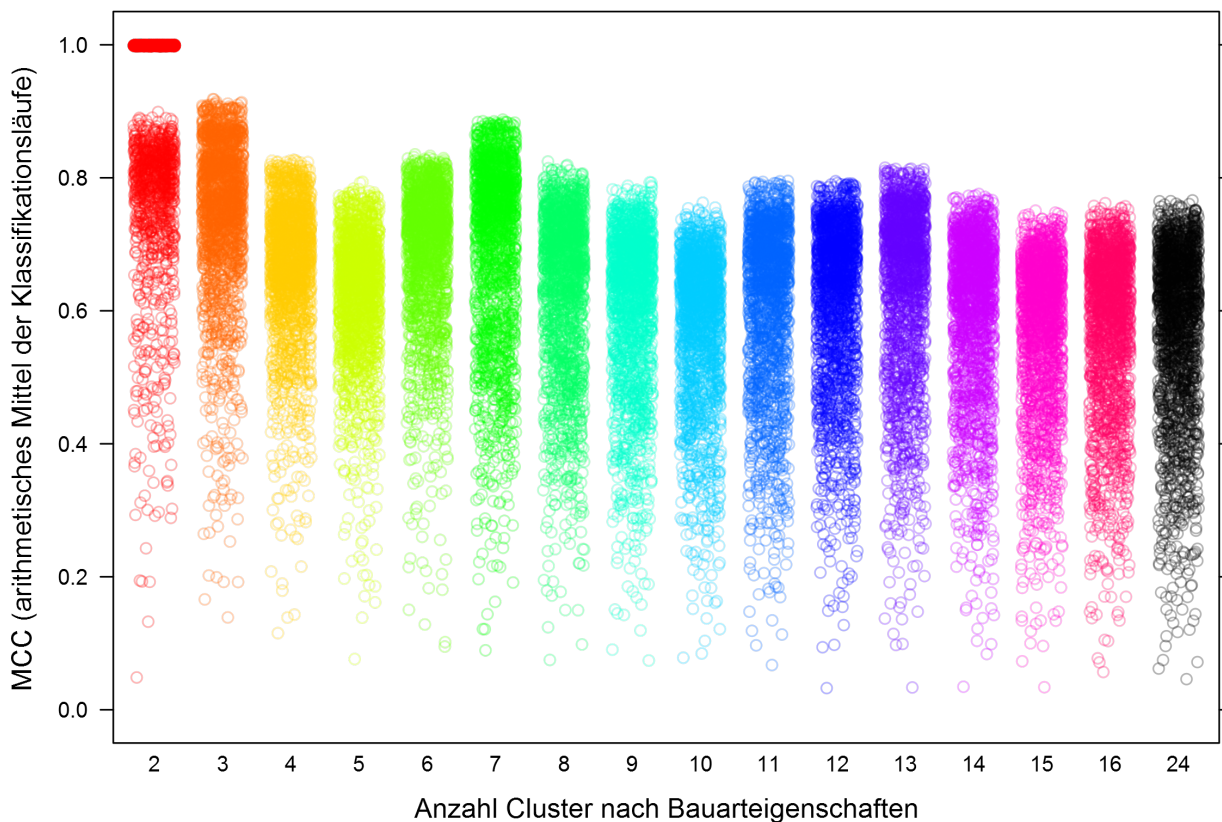


Abbildung 5-11: MCC der Klassifikation der Bauartclusterzugehörigkeit je Clusteranzahl

Die Klassifikationen bei drei und bei sieben Clustern erreichen in Abbildung 5-11 ähnlich gute Ergebnisse. Diese sind sogar besser, als bei zwei Clustern, wenn von den Ausnahmen ($MCC = 1.0$) abgesehen wird. Bei drei Clustern bleibt Weiche 6 eine Gruppe für sich (siehe Abbildung 5-9).

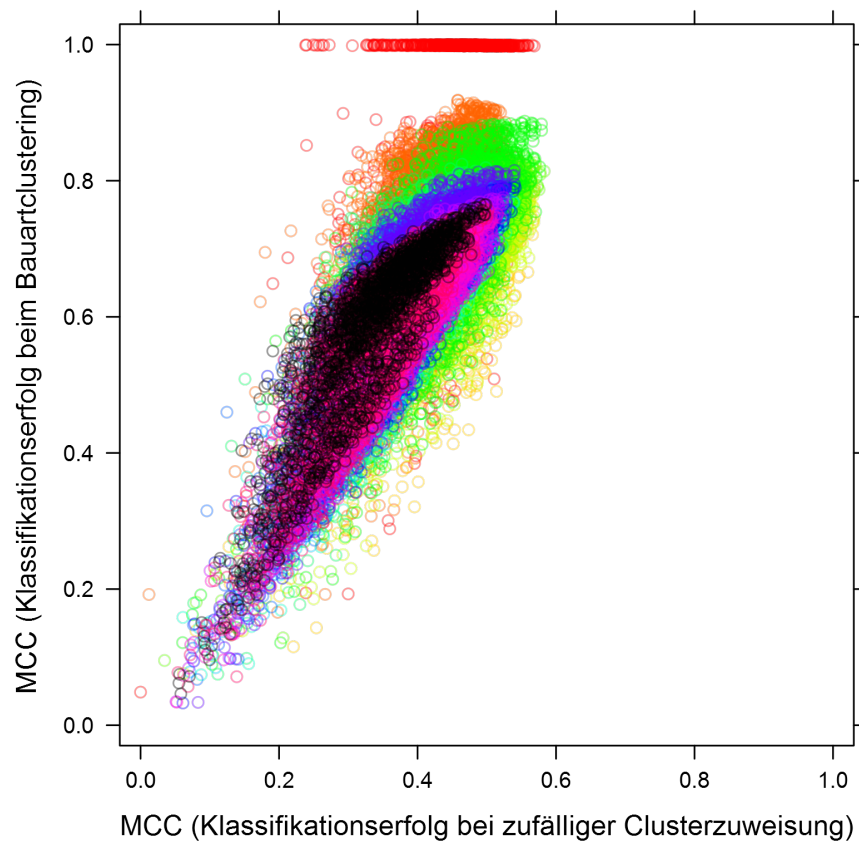


Abbildung 5-12: Gegenüberstellung der MCC-Werte der Zufallscluster- und der Bauartclusterzugehörigkeit (Färbung gemäß Clusteranzahl in Abbildung 5-11)

Die wesentlichen Bauartmerkmale des zweiten Clusters sind der Gabelmittelverschluss und die Gleitstühle ohne Zungenroller. Demgegenüber besitzen die Weichen im dritten Cluster mindestens einen Klinkenmittelverschluss und Zungenroller (vgl. Tabelle 5-7). Eine Aufteilung in weitere Gruppen ist erst wieder bei sieben Clustern sinnvoll, eben wenn die MCC-Werte ähnlich gut ausfallen. Dann wird zusätzlich die Art des Klinkenverschlusses und der Weichentyp relevant. Das wiederum ist konsistent zu einem Cluster für ABW mit Klinkenmittelverschluss und Zungenroller, auch wenn Weiche 6 hier der einzige Vertreter ist. Sieben Cluster bzw. die darin vorliegende Gruppierung nach Weichentyp, Verschluss und Zungenroller erklärt das unterschiedliche Verhalten der Messdaten am besten.

Ein zusätzlich wichtiger Aspekt ist das deutlich ungenauere Klassifikationsergebnis bei 16 oder 24 Clustern. Insbesondere die schlechteren Ergebnisse der weichenindividuellen Klassifikation beweisen, dass die genannten Bauarteigenschaften das Verhalten erklären und nicht die einzelne Weiche selbst. Alle Bauartklassifikationen mit guten Ergebnissen heben sich von der Klassifikation bei zufälliger Gruppierung ab, wie Abbildung 5-12 unterstreicht. Damit ist sichergestellt, dass diese Clusterstrukturen auch signifikant sind. Ferner deckt sich das quantitative Ergebnis bezüglich des Weichentyps, des Verschlusses und der Zungenroller mit den zu Beginn des Abschnitts beschriebenen, qualitativen Aussagen aus der Literatur. Dass eine weitere Gruppenteilung gemäß des Radius und des Winkels, aber vor allem der Abzweigrichtung und der Schwellenart weniger bedeutsam ist, belegen die abnehmenden MCC-Werte bei wachsender Clusteranzahl. Dies steht zwar im Widerspruch zur Literatur, lässt sich jedoch begründen. Die Aussagen zum Einfluss von Radius, Winkel und Schwellenart auf das Verschleißverhalten von Weichen entstam-

men der Arbeit von Zwanenburg [Zwa09], der diesen Merkmalen einen langfristigen Einfluss auf den Zustand der Weiche zuschreibt. Sie sind daher eher für das Ende der Weichenlebensdauer entscheidend als für einzelne Störungen. Zudem hat Zwanenburg den Antrieb von all seinen Untersuchungen ausgeschlossen. Aber gerade der Antrieb verursacht die meisten Störungen, weshalb die Stellstrommessungen eben eine sehr gute Basis für die Prognose darstellen. Daher muss ein Prognosemodell jene Bauarteigenschaften berücksichtigen, die mit dem Antriebsverhalten zusammenhängen.

Zusammengefasst legt die hier durchgeführte Analyse erstmalig offen, welche bauartbedingten Faktoren auf das Verhalten bei der Stellstromüberwachung wirken. Demnach ist während der Entwicklung eines Prognosemodells auf Basis von Stellstrommessungen der Weichentyp, die Verschussart und die Zungenrollvorrichtung zu berücksichtigen. Das ist für die Interpretation der Messwerte wichtig, denn was bei einer Weichengruppe mit gleichen Bauarteigenschaften kritisch ist, ist bei einer anderen Gruppe womöglich unkritisch. Außerdem erreicht das Modell so eine gewisse Allgemeingültigkeit und Übertragbarkeit, da neu zu beurteilende Weichen gemäß ihrer Bauart einsortiert und ihre Messungen entsprechend interpretiert werden können. Bei gänzlich neuen Weichentypen mit neuen Bauarteigenschaften sollte die Analyse wiederholt werden, um auch diese Typen integrieren zu können.

Neben den systematisch untersuchten Einflussfaktoren des Klimas und der Bauart hängt der Zustand von Weichen auch von den Instandhaltungsmaßnahmen ab, die an und um die Weiche herum durchgeführt werden. Ebenso wirken sich die Unterbauqualität, die Liegezeit, die Achslast und die kumulierte Tonnage der Überfahrten auf den Zustand aus. [RB03], [Uni08, S. 16], [Zwa09, S. 81] Mangels verfügbarer Daten können diese Phänomene hier allerdings nicht untersucht werden. Die Erhebung entsprechender Daten wäre nur durch zusätzliche Sensorik möglich gewesen. Um aber die Anforderungen nach minimaler Sensorausrüstung und nach ausschließlicher Verwendung von EBA-zugelassener Messtechnik (siehe Nummer 5 und 6 in Tabelle 4-1) nicht zu verletzen, wird das Modell ohne Berücksichtigung dieser Einflüsse entwickelt.

5.3 Selektion der Rohdatenbasis zur Entwicklung des Prädiktionsmodells

Im bisherigen Verlauf des Kapitels wurden die Messdaten der Stellstromüberwachung als zentrale Basis identifiziert, auf der die Aussagen zum Weichenzustand aufgebaut werden. Allerdings hat die detaillierte Performanceanalyse des originären Weichendiagnosesystems auch gezeigt, dass diese Daten allein nicht die notwendige Genauigkeit liefern können. Weil der Weichenzustand von vielen Faktoren beeinflusst wird, muss in einem ersten Schritt der Prognosemodellentwicklung die Datenbasis erweitert werden. Deshalb wurde hier zum ersten Mal der Zusammenhang von klimabedingten und bauartbedingten Einflussfaktoren systematisch untersucht. Bei den Klimadaten korrelieren die Temperatur und die Luftfeuchtigkeit stark mit einigen Messwerten, weshalb diese beiden Größen Bestandteil der Datenbasis sein sollten. Bei den Bauarteigenschaften sind der Weichentyp, die Verschussart und die Zungenrollvorrichtung entscheidend, um Weichen mit ähnlichem Verhalten zu gruppieren. Im Gegensatz zu Temperatur und Luftfeuchtigkeit sind die drei Bauartmerkmale zwar nicht für die Interpretation jeder Umlaufmessung wichtig, aber für die Einordnung von Weichen gemäß ihres allgemeinen Messwertniveaus.

Von 100 verfügbaren Attributen (A) ergeben sich so 28 Attribute (\mathcal{A}), die als Rohdatenbasis in die Prognosemodellentwicklung einfließen. Aus Platzgründen werden diese hier nur namentlich aufgezählt (ihre Beschreibung kann den Tabellen 5-5, 5-6 und 5-1 oder dem Anhang B.2 entnommen werden):

- ERFID (MessungsID), AKZ (Weichennummer), ErfDatum (Messdatum), Richtung, Pleer, Plauf1, Plauf2, Pverschl, Tumlauf, lunsym, Pzvorspg1, Pzvorspg2, Prutschmax, Umotor, PverschlPeakMax, Luftfeuchtigkeit, Temperatur, StörungsID, StörBegDatum (Störungsbeginn), StörEndDatum (Störungsende), Beschreibung, Ursachentext, ProblCodeTxt (Problemcodetext), ProbGrpText (Problemcodegruppentext), Weichentyp, Verschlussart, Anzahl der Zungenrollpaare

$$\mathcal{A} = \{A_m : 1 \leq m \leq 28\} \subset A \quad \text{mit } |A| = 100 \quad (5.1)$$

Die eben genannten Attribute enthalten auch SAP-Dateneinträge zu Störungen. Diese Störungsdaten dienen als Referenz für den tatsächlichen Zustand (vgl. Abschnitt 5.1.3). Hier muss allerdings eine Einschränkung der genutzten Daten vorgenommen werden. Da die selektierten und auch die verfügbaren Attribute kaum Aussagen zum Zustand von Oberbaukomponenten der Weichen (vgl. Abbildung 2-1) zulassen, sind Störungen mit solchen Ursachen als Referenz zu entfernen. Andernfalls würden sie die Lernphase der Prognosemodellentwicklung negativ beeinflussen. Denn ein Lernalgorithmus würde eventuell Attributwerte bestimmen, die auf solche Störungen hinweisen, obwohl gar kein Zusammenhang besteht. In der Folge käme es zu schwächeren Zuordnungen von Attributwerten zu bestimmbarer Störungen und zu verstärkten Fehlklassifikationen. Das ist ähnlich wie beim menschlichen Lernprozess. Lehrer bringen Schülern eine mathematische Formel durch die Vorgabe der korrekten Formel bei und irritieren sie nicht durch eine falsche.

Die Begründungen, mit welchen 35 Störungen aus der Referenz ausgeschlossen werden, lassen sich folgendermaßen zusammenfassen:

- Insgesamt neun Einträge beschreiben im Grunde Instandhaltungsarbeiten, bei denen etwas eingestellt (Zungenroller, Verschluss) oder eingebaut (Radlenkerbleche) wurde. Dies geschah nach vorheriger manueller Inspektion und konnte nicht aus den Daten abgelesen werden. Dies gilt insbesondere für den Einbau der Radlenkerbleche (vgl. Abbildung 2-3), welche die Zugräder an der Herzstückklücke sicher in der Spur halten. Da hier nachgerüstet wurde, handelt es sich nicht um eine Weichenstörung im eigentlichen Sinne.
- Eine defekte Weichenheizung trat sieben Mal als Störungsursache auf. Obwohl gestörte Heizungen der Weiche zugeordnet werden, kann eine solche Störung nicht mittels der bekannten Parameter überwacht werden. Die Heizungen haben einen eigenen Stromkreis, eine eigene Steuerung und eigene Überwachung. Der eigentliche Zustand der Weiche ist von der Heizung unabhängig.
- In sieben Fällen verursachte eine sogenannte unzeitige Auffahrmeldung die Störung. D.h. die Weiche meldete, dass ihr Verschluss aufgefahren wurde, obwohl gar kein Zug darübergefahren ist. Das passiert, wenn der Endlagenprüfer (ELP) ausrastet. Das wiederum kann vorkommen, wenn der Unterbau in einem schlechten Zustand ist. Dann können im

Nachbargleis vorbeifahrende Züge den Antrieb zum Schwingen bringen, so dass der ELP herausspringt. [Böh13c] Das Phänomen ist über den Stellstrom oder die Weiche selbst nicht zu erfassen.

- Durch eine plötzliche, mechanische Hemmung entstanden vier Störungen, wovon zwei Störungen einmalig auftraten und eine auf Eisabwurf zurückzuführen ist. Solche Ereignisse sind nicht vorhersehbar.
- Drei Mal war eine verbogene Zunge die Ursache für eine Störung. Dieses Phänomen kann nicht mit den gegebenen Daten überwacht werden.
- Gleiches gilt für den losen Herzstückbolzen und den Schienenfehler, der bei der Oberbauinspektion durch eine Ultraschallmessung ermittelt wurde.
- Außerdem werden drei weitere Störungen relegiert. Zwei von ihnen entfallen, da die entsprechende Weiche 6 von der Modellbildung vollständig ausgeklammert wird. Bei dieser Weiche hat SIDIS W rund 78 Prozent seiner Messungen als unvollständig oder fehlerhaft dokumentiert. (Die Werte der anderen Weichen liegen unter 14 Prozent.) Die dritte Störung wird bei Weiche 23 entfernt. Sie ist ein Duplikat, weil sie versehentlich doppelt verknüpft wurde.

Aus Gründen der Vollständigkeit und damit der Leser die Entscheidung nachvollziehen kann, befinden sich in Tabelle B-5 im Anhang B.5 die einzelnen SAP-Einträge und die jeweilige Begründung für den Ausschluss. In Folge des Ausschlusses verbleiben 63 Störungen, die im weiteren Verlauf der Arbeit als Referenz für den tatsächlichen Zustand dienen. Es ist hierbei wichtig zu erwähnen, dass die entsprechenden Anforderungen an die Prognose (Nummer 2 und 3 der Tabelle 4-1) nicht von der Reduktion tangiert werden. D.h. es sind auch weiterhin mindestens 50 Prozent der Ausgangsmenge, also 49 Störungen, rechtzeitig vorherzusagen. Und es sollen dadurch nach wie vor 50 Prozent der VMin reduziert werden.

Angesichts der Reduktion von 35 Störungen stellt sich die Frage: Welchen Einfluss hat dies auf die Performance von SIDIS W? Denn schließlich kann SIDIS W die relegierten Störungen ebensowenig bestimmen, wie ein Prognosemodell. Um die Frage zu beantworten, wurde die Performance Evaluation auf Basis der reduzierten Störungen wiederholt. Das Ergebnis ist jedoch kaum besser als bisher. Die TP Rate verbessert sich zwar ein wenig, aber die FP Rate zeigt keine Verringerung der Fehlalarme (vgl. Abbildung B-4 im Anhang B.5). Eine ausreichende Genauigkeit wird mit dem herkömmlichen Verfahren nach wie vor nicht erreicht, und eine RUL wird von den Diagnosesystemen ja ohnehin nicht bestimmt.

Zur Rohdatenbasis ist abschließend noch zu sagen, dass im Zuge der Selektion und Vorverarbeitung auch die Menge der zu verwendenden Samples (\mathcal{S}) aller Messungen (\mathcal{S}) eingeschränkt wurde, weil:

- für den Beginn der Aufzeichnungsperiode nächtliche Klimadaten fehlen. Bei der Verknüpfung (Inner Join) der SIDIS W Daten mit den Wetterstationsdaten wurden, wegen der Relevanz von Temperatur und Luftfeuchtigkeit, Messungen ohne korrespondierende Klimadaten nicht übernommen.
- Aufzeichnungen deselektiert wurden, wenn SIDIS W diese Messungen als unvollständig oder fehlerhaft identifiziert hat oder die Umstellrichtung nicht erkennen konnte.

- die Messungen der Weiche 6 wegen des hohen Anteils (circa 78 Prozent gegenüber sonst unterhalb 14 Prozent) an unvollständigen oder fehlerhaften Messungen komplett entfernt wurden. Vermutlich war SIDIS W falsch kalibriert oder die Sensorik falsch angeschlossen. Aber solange die Ursache unbekannt ist, werden die Daten vorsichtshalber nicht für die Modellbildung verwendet.

Da jede Messung von SIDIS W interpretiert wird, hat die Samplereduktion keine Auswirkung auf die Performance. Insgesamt bilden demnach 245862 Samples über 28 Attribute für 28 Weichen die Rohdatenbasis für die Modellentwicklung.

$$\mathcal{S} = L \times \mathcal{A} \quad \text{mit } L := \{l \in \mathbb{N} | 1 \leq l \leq |\mathcal{S}|\} \text{ und } |\mathcal{S}| = 245862 \quad (5.2)$$

5.4 Zusammenfassung des Kapitels

Angelehnt an den Prozess der Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD) wurden in diesem Kapitel die Datenquellen und deren Vorverarbeitung beschrieben, die die Grundlage für die datengetriebene Entwicklung des Prognosemodells bilden. Aufgrund des generell hohen Anteils an Störungen, bei denen der Antrieb den Ausfall verursacht, steht dessen Stellstromüberwachung hier im Mittelpunkt. Die Stellstrommessungen stammen von einem für den Bahnbetrieb zugelassenen Weichendiagnosesystem. Der Begründung, hierbei SIDIS W zu nutzen, folgte eine Übersicht über die akquirierte Datenmenge zu 29 Weichen aus einem zweijährigen Zeitraum. Als weitere Datenquelle dienen die dokumentierten Störungen der betreffenden Weichen. Sie bilden die Referenz für den tatsächlichen Weichenzustand, ohne den eine Entwicklung und Evaluation der Prognose nicht möglich wäre. Zunächst wurden die Störungsdaten jedoch verwendet, um die Performance von SIDIS W mittels Receiver Operating Characteristic (ROC)-Metrik zu evaluieren und dadurch offenzulegen, ob die Stellstromwerte und deren Schwellwertalarme bereits für die Modellbildung ausreichen. Doch selbst die verschiedenen Kombinationen der SIDIS W Zustandsalarme (z.B. rote Alarme in x aufeinanderfolgenden Messungen), ein im Rahmen der vorliegenden Forschungsarbeit erstmals untersuchter Ansatz, liefern keine ausreichende Genauigkeit. Wie mit der eigens entwickelten Methode zur Break Even Situation bewiesen werden konnte, ist insbesondere die Genauigkeit aus wirtschaftlicher Sicht nicht gegeben, weil durch Fehlalarme mehr Kosten entstünden, als durch verhinderte Störungen eingespart würden.

Es wurde daraufhin festgestellt, dass zwei Dinge erforderlich sind, um die Genauigkeit zu erhöhen. Zum einen ist die Datengrundlage zu erweitern und zum anderen ist, anstelle der Schwellwerte, eine alternative Zustandsinterpretation zu wählen. Während Letzteres Inhalt des nächsten Kapitels wird, wurde in diesem Kapitel systematisch analysiert, welche weiteren Datenquellen zu berücksichtigen sind. Hierzu wurden erstmals die klimabedingten und die bauartbedingten Zusammenhänge zu Stellstrommessungen quantifiziert bzw. identifiziert. Aus einer Reihe von Klimadaten zeigten die Temperatur und die Luftfeuchtigkeit eine hohe Korrelation mit den Stellstrommessungen. Dies ist ein klarer Beweis dafür, dass beide Größen bei der stellstrombasierten Prognose zu berücksichtigen sind. Des Weiteren wurde untersucht, wie die zum Teil erheblichen Unterschiede zwischen den Stellstromwerten der Weichen durch Bauarteigenschaften zu erklären

sind. Der Zusammenhang wurde mit dem neuartigen heuristischen Analyseverfahren nachgewiesen, bei dem alle möglichen Cluster mit Weichen ähnlicher Bauartigenschaften gebildet wurden. Die anschließende Klassifikation der Bauartclusterzugehörigkeit anhand der Stellstromwerte erreichte dabei die besten Werte, wenn die Weichen gemäß des Weichentyps, der Verschlussart und der Zungenrollvorrichtung gruppiert wurden. Im Ergebnis werden von diversen Bauartigenschaften nur diese drei genutzt, um Weichen zu gruppieren. Statt Weichen individuell zu behandeln, kann die Prognose auf diese Gruppen ausgerichtet werden, weil Weichen innerhalb der Gruppen ähnliche Stellstromverläufe besitzen.

Das Kapitel schließt mit einer kurzen Erläuterung der Rohdatenbasis für die Prognosemodellentwicklung. Diese besteht nunmehr aus rund 245000 Samples über 28 Attribute zu Stellströmen, Klima, Bauart und Störungen bei 28 Weichen. Es wurde dabei insbesondere erläutert, dass einige Störungen nicht in die Modellbildung eingehen, weil sie nicht über die Rohdatenbasis interpretierbar sind. Zudem erfüllen die Rohdaten bereits einen Teil der Anforderungen, die in Kapitel 4 formuliert wurden, wie die Tabelle 5-9 zeigt.

Nummer	Kurzbeschreibung	Absicht	Status
1.	Vorhersage in Form einer verbleibenden Nutzungsdauer (engl. Remaining Useful Life) (RUL)	Planbarkeit	?
2.	Vorhersage mit richtig Positiv (TP) Rate größer 0.5 für die Klasse der RUL von fünf oder weniger Tagen	mehr Pünktlichkeit in Deutschland (DE)	?
3.	Summe der Verspätungsminuten (VMin) von TP in der Klasse mit der RUL von fünf oder weniger Tagen ist größer 50%	mehr Pünktlichkeit in DE	?
4.	Erreichen der Break Even Situation	Wirtschaftlichkeit	?
5.	Einsatz von so wenig Sensoren wie möglich	Wirtschaftlichkeit	✓
6.	Nutzung von Sensoren, die für den regulären Bahnbetrieb zugelassen sind	Wirtschaftlichkeit	✓
7.	Verwendung von Weichen mit verschiedenen Eigenschaften	Allgemeingültigkeit	✓
8.	Verwendung von mindestens zehn Weichen	Allgemeingültigkeit	✓
9.	Verwendung von häufig im Netz der DB AG verbauten Weichen	Anwendbarkeit in DE	✓
10.	Erhebung von realen Mess- und Störungsdaten	Anwendbarkeit in DE	✓
11.	Datenbasis über den Zeitraum von mindestens einem Jahr	Anwendbarkeit in DE	✓

Tabelle 5-9: Status (✓ = erfüllt, ? = offen, x = nicht erfüllt) aller Anforderungen an die zu entwickelnde Prognose mit der Rohdatenbasis

Bevor im nächsten Kapitel das eigentliche Prognosemodell beschrieben wird, sei hier noch ein Punkt in Bezug auf die Dissertation von Asada [Asa13] hervorgehoben. Die erfüllten Anforderungen decken alle Vorschläge für zukünftige Forschungsarbeiten ab (reale Felddaten, langer Zeitraum, verschiedene Weichentypen, wechselnde Umweltbedingungen), die von Asada genannt werden [Asa13, S. 143 ff.]. Da seine Publikation eine der jüngsten zur Weichendiagnose ist, macht dies deutlich, dass die vorliegende Arbeit nicht nur durch die Prognose neues Terrain betritt, sondern auch beim Datenumfang und der Realitätsnähe.

6 Prädiktionsmodell für den Zustand von Eisenbahnweichen

Die ersten Kapitel erläuterten grundlegend die Funktion und den Aufbau von Weichen (Kapitel 2) sowie Verfahren der Zustandsprädiktion und Datenanalyse samt passender Performance Metriken (Kapitel 3). Die Analyse bestehender Ansätze für die Weichenzustandsbeurteilung ergab, dass vor allem bisher keine zuverlässige Prognose der verbleibenden Nutzungsdauer (RUL) existiert. Deshalb wurde dies ein essentieller Punkt der in Kapitel 4 aufgestellten Anforderungen an das zu entwickelnde Prognosemodell. Wie Kapitel 5 anschließend zeigte, sind dessen zentrale Basis Stellstrommessungen, die um Daten zum Klima und zu den Bauarteieigenschaften ergänzt werden. Der Umfang und die Heterogenität der so zusammengestellten Rohdatenbasis unterstreichen noch einmal die Richtigkeit, sich bei der datengetriebenen Entwicklung am systematischen Prozess der Wissensentdeckung in Datenbanken (KDD) zu orientieren.

Diesem Prozess folgend, wird in diesem Kapitel das Prädiktionsmodell mit Hilfe von Data Mining Verfahren erstellt und mit Performance Metriken bewertet. Dazu werden zunächst die fundamentale Idee für die Prognose beschrieben und – ausgehend von den Eigenschaften der Datenbasis – passende Prognoseverfahren und Metriken ausgewählt oder erstellt (Abschnitt 6.1). Anschließend werden die Parameter und Einstellungen der verwendeten Prognoseverfahren dahingehend untersucht, dass eine maximale Performance erreicht wird. Dies ist ein wichtiger Schritt im iterativen KDD-Prozess, da die Ergebnisqualität einiger Verfahren stark von den gewählten Parametern abhängt. Daraufhin werden die individuell optimal parametrisierten Verfahren miteinander verglichen, wenn diese unter bestimmten Bedingungen angewendet werden. Im Ergebnis zeigt sich, inwieweit die Prognose die gestellten Anforderungen erfüllt, insbesondere die Anforderung nach dem Erreichen der Break Even Situation (Abschnitt 6.2). Es sollen schließlich nicht nur Störungen verhindert, sondern auch Instandhaltungskosten gesenkt werden. Bis dahin werden Daten aus dem gesamten Zeitraum verwendet, um die Prognose zu entwickeln. Denn es ist zunächst wichtiger, möglichst viele Phänomene des Abnutzungsverhaltens und der Störungsentwicklung einzubinden. Die Rohdatenbasis beinhaltet ja lediglich 63 Störungen, die zudem aus verschiedenen Ursachen resultieren können. Eine möglichst hohe Anzahl an Referenzen wird anfangs benötigt, um überhaupt die Möglichkeiten der Prognose zu ergründen. Die Anwendung der entwickelten Prognose unter realen Bedingungen wird daher zum Abschluss des Kapitels betrachtet (Abschnitt 6.3).

6.1 Fundamentales Konzept für die Prognose

Die detaillierte Analyse der klassischen Weichendiagnose mit ihren Alarmschwellwerten führte zu dem Schluss, dass einerseits die Datenbasis erweitert werden musste und andererseits eine alternative Methode zur Zustandsbeurteilung zu finden ist (vgl. Abschnitt 5.2). Insbesondere die hochdimensionalen, heterogenen, nicht normalverteilten Daten mit nicht linearen Zusammenhängen verlangen leistungsfähige Verfahren für eine Prognose. Der folgende Abschnitt beschreibt das grundlegende Konzept, mit dem die RUL der Weichen vorhergesagt werden soll. Im Wesentlichen teilt sich das Konzept in zwei Teile auf:

1. Definieren von Klassen der RUL, die nicht zu detailliert sind, aber auch eine gute Planbarkeit für die Instandhaltung sicherstellen. Dadurch wird die Prognose zu einer Klassifikationsaufgabe multipler Klassen.
2. Erlernen und Optimieren der Klassenzugehörigkeit anhand der Rohdatenbasis, so dass die Klassifikationsgenauigkeit den Anforderungen genügt.

6.1.1 Transformation der stetigen Restnutzungsdauer in multiple Klassen

Der erste Teil des fundamentalen Konzepts für die Zustandsprognose greift einen Ansatz dieser Arbeit wieder auf, der im Abschnitt 4.2.1 eingeführt wurde: die Transformation der stetigen RUL in diskrete Klassen. Wie in dem Abschnitt bereits begründet wurde, ist eine stetige RUL bei Weichen nicht sinnvoll:

- Zum einen würde damit eine Granularität und Präzision der Prognose vermittelt, die tatsächlich stark schwanken kann, wie der erfolglose Versuch der Störungsvorhersage mittels linearer Regression in Böhm [Böh12c] unterstreicht. Die Randbedingungen schwanken über einen längeren Zeitraum stärker, z.B. wegen wechselnder Belastungen oder Instandhaltungsmaßnahmen des Oberbaus. Je weiter in der Zukunft die Störung liegt, desto unsicherer ist die Prognose erwartungsgemäß. Beispielsweise träfe die Störung tatsächlich in 90 Tagen ein und würde heute mit 87 Tagen vorhergesagt, morgen aber mit 92, übermorgen mit 90 und tags darauf mit 85 Tagen. Es könnte beim Nutzer der Eindruck eines unzuverlässigen Systems entstehen. Außerdem stellt sich die Frage, ob Nutzer einem System vertrauen schenken, dass die RUL stunden- oder minutengenau angibt, während die Störung mehrere Tage, Wochen oder gar Monate in der Zukunft liegt.
- Zum anderen sind auch die Zustandsmessungen nicht stetig oder besitzen äquidistante Intervalle, da sich Umlaufzeitpunkt und -frequenz bei Weichen unterscheiden.
- Hinzu kommt, dass in der Praxis die Instandhaltungseinsätze über den aktuellen Monat hinaus nicht im Stundenrhythmus geplant werden, sondern als Schicht oder Schichtwoche.

Eine diskrete Einteilung der RUL ist demnach vorteilhaft. Doch während sich die Transformation in Abschnitt 4.2.1 lediglich auf die binäre Einteilung in nicht rechtzeitig und rechtzeitig erkannte Störungen bezog, sind nun mehr Klassen von Interesse. Wie viele Klassen das sein sollten und wie diese Klassen aufzuteilen sind, ist schwer einzuschätzen, weil die Bedürfnisse oder die Erfahrungswerte einer Instandhaltung mittels Zustandsprognose unbekannt sind – wie auch, wenn es ja bisher keine Vorhersage gibt.

Den einzigen Anhaltspunkt bilden die bekannten Vorlaufzeiten für die Notfallentstörung (ein bis zwei Stunden vor Störungseintritt) und für eine in die Feinplanung einzuarbeitende Instandhaltungsmaßnahme (drei bis fünf Tage) [Böh13c]. Beide sollten bei einer Diskretisierung berücksichtigt werden. Des Weiteren soll bei der Einteilung bedacht werden, dass die Zeitintervalle für Klassen umso weiter gefasst werden, je mehr Zeit bis zur Störung bleibt. Dadurch werden die oben erwähnten Schwankungen in der RUL-Prognose vermieden, was letztlich das Vertrauen des Nutzers in die Vorhersage stärkt. Außerdem ist aus implementierungstechnischen Gründen eine Einteilung wünschenswert, die sich mathematisch gut beschreiben lässt. Gibt es eine regelmäßige endliche Folge für die Transformation, die diese drei Punkte erfüllt und wenn ja welche?

Die Antwort bildet die geometrische Zahlenfolge der Zweierpotenzen (2^x). Sie ist nicht nur mathematisch eindeutig definiert und leicht zu berechnen, sondern passt auch gut zu den Zeitgrößen, wenn der Wert der Zweierpotenzen als Tage (d) gewertet wird. So ergeben z.B. $2^{-4} = 0.0625$ Tage bzw. 1.5 Stunden und $2^2 = 4$ Tage. Beide repräsentieren jeweils genau die Mitte der Vorlaufzeiten für die Notfallentstörung und die Feinplanung.

Die Gleichung (4.7), welche für die Transformation der stetigen RUL in binäre Klassen verwendet wurde, wird daher mit Hilfe der Zweierpotenzen folgendermaßen verändert:

$$Z: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 14\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq 2^{-4} \\ i: 2^{i-6} & < RUL \leq 2^{i-5} \\ 14: 2^8 & < RUL \end{cases} \quad (6.1)$$

$$TZ: \mathcal{S} \rightarrow Z$$

$$s_l \mapsto z_i \quad \forall s_l \in \mathcal{S}: A_{\text{ZeitbisStörung}}(s_l) \neq \emptyset \quad (6.2)$$

Da der Zeitpunkt aller Störungen bekannt ist, wird für jedes Sample die Zeit bis zur Störung berechnet und als Attribut ($A_{\text{ZeitbisStörung}}$) hinzugefügt. Wenn bei einer Weiche keine Folgestörung mehr enthalten ist, bleibt das Attribut leer (in der Statistik bekannt als rechtszensierte Daten des Typs 3). Die stetige RUL jeder Messung wird durch die Gleichung (6.1) einer der 14 Zustandsklassen zugeordnet. Diese Anzahl wurde gewählt, weil zum einen die Instandhaltung bei weniger als 1.5 Stunden nicht mehr reagieren kann, und zum anderen weil Klassen über 256 Tagen kaum noch planungsrelevant erscheinen. Die Tabelle 6-1 listet die entstehenden Klassen (z_i) und das RUL-Intervall auf, welches die Klasse jeweils abdeckt.

Klasse z_i	Intervall der RUL [d]	Anzahl Samples
1	$0 \leq RUL \leq 0.0625$	248
2	$0.0625 < RUL \leq 0.125$	148
3	$0.125 < RUL \leq 0.25$	266
4	$0.25 < RUL \leq 0.5$	434
5	$0.5 < RUL \leq 1$	996
6	$1 < RUL \leq 2$	1933
7	$2 < RUL \leq 4$	3801
8	$4 < RUL \leq 8$	6888
9	$8 < RUL \leq 16$	13435
10	$16 < RUL \leq 32$	22439
11	$32 < RUL \leq 64$	25307
12	$64 < RUL \leq 128$	33310
13	$128 < RUL \leq 256$	31008
14	$256 < RUL \leq \infty$	44065

Tabelle 6-1: Klassen der RUL mit ihren jeweiligen Intervallen in Tagen

Die getroffene Einteilung verspricht auch für die operative Anwendung einige Vorteile, weil sie kalendarischen Zeiteinteilungen ähnelt. Bspw. könnte ein Tag Vorlauf genügen, um die Instand-

haltungseinsätze für den Folgetag umzudisponieren. Acht Tage bedeuten annähernd eine Woche. Und 32 bzw. 64 Tage entsprechen etwa einem bzw. zwei Monaten. Natürlich sind auch andere Einteilungen denkbar. Doch da zum jetzigen Zeitpunkt eben keine besseren Gründe für eine andere Transformationsvorschrift existieren, wird die Klasseneinteilung wie beschrieben vorgenommen. Im Verlauf des Kapitels werden andere Einteilungen untersucht, um zu zeigen, inwieweit sich dadurch das Klassifikationsergebnis ändert.

Das in diesem Abschnitt definierte Vorgehen wandelt die Vorhersage in ein Klassifikationsproblem multipler Klassen um. Damit unterscheidet sich der gewählte Prognoseansatz von den meisten gängigen RUL-Vorhersagen. Die Prognose vermeidet dadurch nicht nur überpräzise stetige RUL-Angaben, sondern verzichtet auch auf aufwendige oder unsichere Extrapolationen von Zeitreihen. Es gilt nun den zweiten Teil des grundlegenden Konzepts vorzubereiten, nämlich die Auswahl von Klassifikationsverfahren und Metriken, welche zur Daten- und Problemstruktur passen.

6.1.2 Auswahl von Klassifikationsverfahren für die Prognose der Restnutzungsdauer und Metriken zu deren Bewertung

Nachdem die Zustandsklassen (I) gebildet wurden, können nun passende Verfahren und Metriken ausgewählt werden. Denn jetzt steht, zusammen mit der Rohdatenbasis, die Struktur des Problems für die Prognosemodellbildung fest. Alle Samples (\mathcal{S}) sind anhand der Attribute (\mathcal{A}) einem Zustand (Z) zuzuordnen bzw. die einzelnen Messungen (s_l) einer Zustandsklasse (z_i).

$$\begin{aligned} f: \mathcal{S} &\rightarrow Z \\ s_l &\mapsto z_i \quad \forall s_l \in \mathcal{S}_{\text{Training}} \subset \mathcal{S} \end{aligned} \quad (6.3)$$

$$\begin{aligned} HZ: \mathcal{S} &\rightarrow Z \\ s_l &\mapsto z_j \quad \forall s_l \in \mathcal{S}_{\text{Test}} = \mathcal{S} \setminus \mathcal{S}_{\text{Training}} \wedge A_{\text{ZeitbisStörung}}(s_l) \neq \emptyset \end{aligned} \quad (6.4)$$

$$\mathbf{C} \in \mathcal{M}(I \times I, \mathbb{N}) \quad \text{mit } C_{i,j} = |\{s \in \mathcal{S} : TZ(s_l) = z_i, HZ(s_l) = z_j\}| \quad (6.5)$$

Die Funktion f ist zunächst anhand der Trainingsdaten zu erlernen und dann auf die Testdaten anzuwenden. Dabei entsteht der hypothetische Zustand (HZ), welcher in der Konfusionsmatrix (\mathbf{C}) dem tatsächlichen Zustand (TZ) aus der Gleichung (6.2) gegenübergestellt wird. Im Abschnitt 3.2 wurden einige Verfahren für diese Prädiktionsaufgabe beschrieben. Da es kein ultimativ bestes Verfahren gibt, sondern jedes Vor- und Nachteile aufweist, hängt der Erfolg einer Methode u.a. von der zugrundeliegenden Problem- und Datenstruktur ab. Im vorliegenden Fall weisen die Prognoseaufgabe und die Rohdatenbasis folgende relevante Eigenschaften auf:

- Es sind relativ viele Klassen zu bestimmen, deren Anzahl ggf. variiert (einige Weichen besitzen keine weit entfernten Störungen). Damit die Prognoseergebnisse stabil sind, sollten Verfahren eine eher geringe Sensitivität gegenüber der Klassenanzahl besitzen.
- Die Messdaten sind stark verrauscht. Verfahren müssen demgegenüber robust sein.
- Wie Tabelle 6-1 zeigt, ist die Anzahl an Samples in den Klassen extrem unbalanciert, so

dass die Prädiktionsmethoden sich dafür eignen müssen.

- Die Anzahl der relevanten Attribute ist mit 13 zwar nicht sonderlich hoch aber auch nicht niedrig, weswegen eher Verfahren für hochdimensionale Probleme zu präferieren sind.
- Weil das Messverhalten von Weichen unterschiedlich ist, sollten die Methoden nicht zur Überanpassung neigen. Nicht zuletzt wird dadurch auch die Allgemeingültigkeit des Prognosemodells unterstützt.

Ein Vergleich mit den in Tabelle 3-1 zusammengefassten Eigenschaften der Prädiktionsmethoden zeigt, dass sich grundsätzlich vier davon eignen:

- k-nächsten-Nachbarn (engl. k-Nearest-Neighbor) (kNN),
- künstliches Neuronales Netz (engl. artificial Neural Network) (aNN),
- Entscheidungsbaum (engl. Decision Tree) (DTree) und
- Stützvektormaschine (engl. Support Vector Machine) (SVM)

Im Einzelnen kann die Performance der vier Verfahren recht unterschiedlich ausfallen. Dies gilt insbesondere für aNN und SVM, weil sie auf zahlreiche Weise parametrierbar werden können (vgl. Abschnitt 3.2.3 und 3.2.5). Um die Prognoseergebnisse evaluieren zu können, sind auch passende Metriken zu verwenden. Es sei noch einmal daran erinnert, dass die richtige Metrik essentiell ist, denn die Wahl der Metrik entscheidet über die Wahl des Modells. Performance Metriken, wie sie in Abschnitt 3.3 grundlegend erläutert wurden, müssen dabei folgende Merkmale besitzen:

- Sie müssen sowohl für binäre als auch für multiple Klassen anwendbar sein, weil die Break Even Situation auf zwei Klassen beruht, während die eigentliche Prognose multiple Klassen bestimmt.
- Die Bewertung sollte möglichst invariant gegenüber einer veränderten Klassenverteilung sein, weil bei wiederholtem Klassifikationstraining und -test die jeweils zufällig sortierte Menge variieren kann.
- Die Metriken müssen bei Rauschen in den tatsächlichen Klassen und den Klassenzugehörigkeitswerten geeignet sein, zumindest bis zu einem gewissen Grad an Rauschen.
- Da die besonders wichtigen Klassen kurz vor Störungseintritt wenig Samples enthalten, sollten Metriken diese Klassen mindestens genauso gewichten, wie Klassen mit vielen Samples.
- Aus dem gleichen Grund sollten bei der Evaluation auch kleine Änderungen angezeigt werden, vor allem wenn es viele Klassen gibt.

Werden diese Merkmale mit den Eigenschaften der Metriken in Tabelle 3-8 abgeglichen, ist festzustellen, dass keine Metrik alle Eigenschaften bestmöglich in sich vereint. Nur der Receiver Operating Characteristic (ROC)-Graph und die makroskopische Fläche unter der ROC-Kurve (engl. Area Under the ROC Curve) (macroAUC) erfüllen die Kriterien, hauptsächlich wegen der Invarianz gegenüber der Klassenverteilung. Während beide Metriken kleine Klassifikationsunterschiede nur mittelmäßig ausweisen, sind beim ROC-Graph zusätzlich Abstriche bei wichtigen aber seltenen Klassen zu machen. Dafür ist der ROC-Graph bei einer detaillierten Betrachtung

vorteilhafter, weil er keine singuläre Größe ist, sondern aus falsch Positiv (FP) Rate und richtig Positiv (TP) Rate besteht. Wenn, wie hier, kleine Unterschiede im Klassifikationsergebnis bei vielen Klassen darzustellen sind, kommt der Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient) (MCC) den beiden anderen Metriken am nächsten. Aus diesem Grund wurde er bereits bei der Bauartklassifikation verwendet (siehe Abschnitt 5.2.3.3). Wegen ihrer unterschiedlichen Stärken, werden in der Arbeit, je nach Evaluationsziel, folgende Metriken genutzt:

- Der macroROC-Graph wird aus der Konfusionsmatrix (**C**) gebildet. Während bei binären Klassen immer eine positive und eine negative Klasse für die Bildung von FP, TN, TP und FN definiert ist, kann dies in der allgemeinen Form der ROC-Graph-Erweiterung auf multiple Klassen nach Hand und Till [HT01] nicht mehr unterschieden werden. Statt eines einfachen ja-nein-Problems (z.B. ob eine Störung vorliegt), wird eine differenziertere Aussage getroffen. Im paarweisen Vergleich der Klassen wird deshalb eigentlich jede Klasse einmal als positive und einmal als negative Klasse behandelt. Im vorliegenden Fall ist jedoch die Klasse mit der kleineren RUL immer die positive Klasse, da sie die handlungsrelevante Klasse ist. Um eine Störung zu vermeiden, reagiert die Instandhaltung immer auf die kürzeste Prognose. Deswegen wird die Berechnung der FP und TP Rate entsprechend adaptiert:

$$macroFP_r = \frac{2}{I(I-1)} \sum_{i=1}^I \sum_{\substack{j=1 \\ i < j}}^I \frac{C_{ji}}{C_{ji} + C_{jj}} \quad (6.6)$$

$$macroTP_r = \frac{2}{I(I-1)} \sum_{i=1}^I \sum_{\substack{j=1 \\ i < j}}^I \frac{C_{ii}}{C_{ii} + C_{ij}} \quad (6.7)$$

- Die macroAUC ist eine singuläre Metrik mit Invarianz gegenüber der Klassenverteilung, die angibt, wie gut ein Klassifikator grundsätzlich in der Lage ist, die Klassen voneinander zu unterscheiden. Wegen der definierten positiven Klassen wird hier ebenfalls die Grundform aus Hand und Till [HT01] im paarweisen Vergleich angepasst:

$$macroAUC = \frac{1}{I(I-1)} \sum_{i=1}^I \sum_{\substack{j=1 \\ i \neq j}}^I \frac{1}{2} \int \hat{F}(\hat{p}|z_i) \hat{f}(\hat{p}|z_j) d\hat{p} = \frac{1}{I(I-1)} \sum_{i=1}^I \sum_{\substack{j=1 \\ i \neq j}}^I \frac{1}{2} AUC(z_i, z_j) \quad (6.8)$$

$\hat{p} :=$ geschätzte Wahrscheinlichkeit, dass ein Sample s_l der Klasse z_i angehört
 $\hat{f}(\hat{p}|z_i) :=$ aus **C** geschätzte Wahrscheinlichkeitsfunktion für \hat{p} ,
 wenn s_l tatsächlich aus der Klasse z_i stammt
 $\hat{F}(\hat{p}|z_i) :=$ aus **C** geschätzte, kumulierte Verteilungsfunktion von \hat{f}

Dieser Berechnung wird eine weitere Besonderheit der Weichenzustandsprognose hinzugefügt. Bereits die optische Betrachtung der Stellstromentwicklung über einen längeren Zeitraum (z.B. in Abbildung 5-6) hat gezeigt, dass sich Werte kurz vor der Störung deutlich

von Werten nach der Entstörung unterscheiden können. Es sollte also für jeden Klassifikator einfacher sein, Klassen zu unterscheiden, deren RUL weit auseinander liegt. Demzufolge sind Klassifikatoren zu bevorzugen, die ähnliche, sozusagen benachbarte Klassen besser identifizieren können als andere Klassifikatoren. Bspw. sei angenommen, dass zwei Klassifikatoren A und B eine ähnliche Performance erreichen. Wenn A die Klassen z_3 und z_4 genauer bestimmt als der Klassifikator B, aber B treffsicherer z_3 von z_9 unterscheidet, dann sollte A bevorzugt werden. Das ist gerade in Bezug auf die Intervallgrenzen der Klassen bedeutsam, da hier die eigentlich fließenden Übergänge diskretisiert wurden und die Vorhersage möglichst wenig schwanken sollte. Es wird daher ein Faktor für die macroAUC eingeführt, die das paarweise Ergebnis umso stärker gewichtet, je dichter die Klassen beieinander liegen. Die ähnlichkeitsgewichtete, makroskopische Fläche unter der ROC-Kurve (engl. weighted, macroscopic Area Under the ROC Curve) (wAUC) wird bestimmt durch:

$$wAUC = \frac{1}{\frac{I}{3}(2I^2 - 3I + 1)} \sum_{i=1}^I \sum_{\substack{j=1 \\ i \neq j}}^I \frac{1}{2} AUC(z_i, z_j) (I - |j - i|) \quad (6.9)$$

- Als dritte Metrik fällt die Wahl auf den Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient) (MCC). Er wird zwar angewendet, wie von Matthews [Mat75] eingeführt, aber der Vollständigkeit halber wird seine Formel ebenfalls angegeben:

$$MCC = \frac{\sum_{i,j,o=1}^I C_{ii}C_{oj} - C_{ji}C_{io}}{\sqrt{\sum_{i=1}^I \left(\sum_{j=1}^I C_{ji} \right) \left(\sum_{\substack{p,q=1 \\ p \neq i}}^I C_{qp} \right)}} \sqrt{\sum_{i=1}^I \left(\sum_{j=1}^I C_{ij} \right) \left(\sum_{\substack{p,q=1 \\ p \neq i}}^I C_{pq} \right)} \quad (6.10)$$

Die drei Metriken werden im weiteren Verlauf verwendet, um die vier Klassifikationsverfahren zu evaluieren, mit denen der Weichenzustand prognostiziert wird.

6.2 Experimenteller Vergleich der ausgewählten Klassifikationsverfahren

Am Anfang des Kapitels wurde das grundlegende Konzept für die Weichenzustandsprognose erläutert. Diesem zufolge wurde die RUL in wohldefinierte Klassen diskretisiert. Dadurch soll es möglich werden, die Klassenzugehörigkeit anhand der Attribute zu erlernen und anschließend Messungen einer bestimmten RUL-Klasse zuzuordnen. Die Diskretisierung ist zwar in erster Linie die Basis für die Zustandsprädiktion, aber auch Teil des Datentransformationsschritts im Prozess der KDD (siehe Abschnitt 3.1), an dem sich diese Arbeit orientiert. Um die Klassifikation vorzubereiten, werden die SIDIS W Messdaten und die Klimadaten im Intervall $[0, 1]$ normalisiert (Min-Max-Normalisierung), damit diese gleich skaliert sind. Auf diese Weise wird verhindert, dass einzelne Attribute die Ergebnisse allein durch große Werte dominieren, z.B. bei der euklidischen Distanzberechnung. Da es sich bei der Normalisierung um ein Standardverfahren handelt, sei diese hier nur zur besseren Nachvollziehbarkeit erwähnt.

Für die Klassifikation der RUL wurden vier potenzielle Verfahren ausgewählt. Ein erster Prognoseversuch, durchgeführt um die grundsätzlichen Erfolgschancen einschätzen zu können, liefert nur mittelmäßige Ergebnisse. Wie Abbildung 6-1 zeigt, verhalten sich die Verfahren recht unterschiedlich. Abgesehen von der SVM erzielen alle eine kleine Fehlalarmrate ($macroFP_r$). Doch während das aNN kaum Klassen richtig identifiziert, klassifizieren kNN, DTree und SVM die RUL zu 50 bis 66 Prozent richtig. DTree und kNN können die Klassen zudem einigermaßen gut voneinander trennen, die verbleibenden Verfahren eher nicht (siehe Abbildung 6-1[b]). Insgesamt deuten die Ergebnisse zwar in eine positive Richtung, reichen aber noch nicht für eine verlässliche Prognose. Doch wie lassen sich diese Resultate erklären?

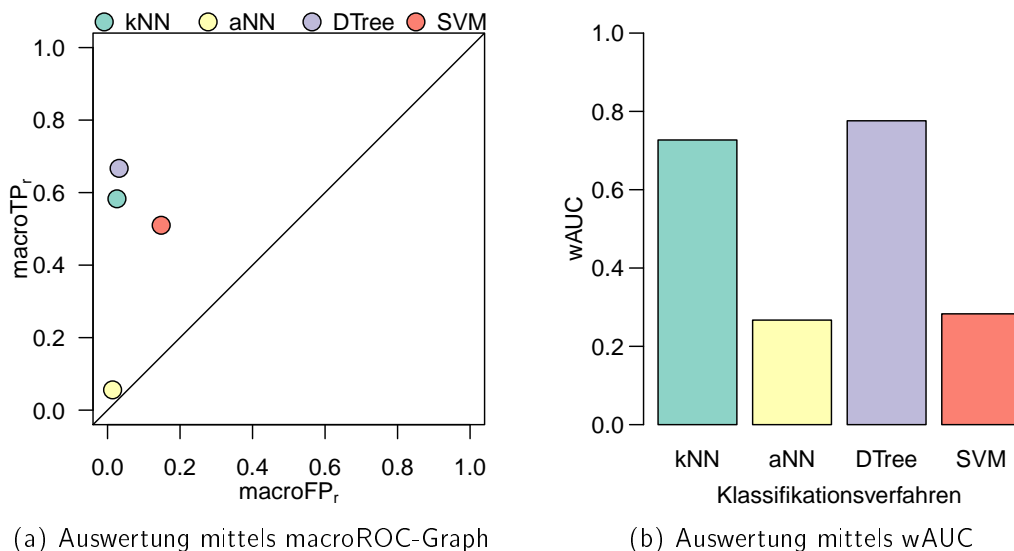


Abbildung 6-1: Ergebnis eines ersten Prognoseversuchs mit zufällig gewählten Parametern für die vier Klassifikationsverfahren

Die Performance kann eine Folge der zufällig gewählten Samples für Training und Test sein. So könnten in beiden Mengen jene Daten enthalten sein, die nicht gut zu lernen oder zu klassifizieren sind. Um im ersten Versuch Rechenzeit einzusparen, wurde mit der halben Samplemenge gearbeitet, so dass sich dieser Effekt einer ungünstigen Auswahl noch verstärkt haben könnte. Die Hauptursache liegt vermutlich in den Einstellungen der Verfahren. Diese wurden im ersten Versuch willkürlich gesetzt, da gute Einstellungen in der Regel nicht im Vorhinein bekannt sind. Wie wiederholt erwähnt wurde, hängt jedoch die Performance der Verfahren stark davon ab, wie sie parametrisiert werden. Dies gilt insbesondere für das aNN und die SVM. Bevor also die Klassenstruktur, die Attribute oder gar der Prognoseansatz verändert werden, sollten die vier Verfahren zunächst individuell betrachtet werden, um jeweils die besten Einstellungen zu bestimmen. Es ist dabei nicht ratsam, lediglich kNN und DTree weiterzuverfolgen, denn beide besitzen Eigenschaften, die bei der operativen Betriebsanwendung nachteilig sein können (vgl. Tabelle 3-1). So ist der Speicherbedarf für kNN sehr hoch und wächst permanent, da die Ähnlichkeit zu historischen Daten immer direkt bestimmt wird. DTree neigen eher zur Überanpassung und sind aufwendig zu implementieren. Sowohl aNN als auch SVM heben einige der Nachteile auf, weshalb im Folgenden für alle vier Verfahren die Parametrierungsoptionen analysiert werden (Abschnitt 6.2.1). Danach wird der Prognoseerfolg bei optimalen Einstellungen erneut verglichen (Abschnitt 6.2.2) und anschließend der Effekt alternativer RUL-Klassen untersucht (Abschnitt 6.2.3). Beides fließt in die Prognoseevaluation aus ökonomischer Sicht ein (Abschnitt 6.2.4).

6.2.1 Individuelle Maximierung des Prognoseerfolgs der Klassifikationsverfahren anhand ihrer Parametrierung

Bei welchen Einstellungen erreichen die vier ausgewählten Klassifikationsverfahren jeweils ihren maximalen Prognoseerfolg? Die systematische Suche nach passenden Verfahren und ihren bestmöglichen Einstellungen ist enorm zeitaufwendig, aber ein wichtiger Forschungsbeitrag für die Lösung unzähliger Klassifikationsprobleme in der Biologie, der Medizin, der Wirtschaft, etc. Das gilt insbesondere für die vorliegende Arbeit zur Weichenzustandsprognose, weil das dafür notwendige mathematische, daten- und implementierungstechnische Know-how nicht Schwerpunkt eines Infrastrukturbetreibers ist. Auch im KDD-Prozess ist die richtige Parametrierung ein essentieller Bestandteil für das Data Mining. Dieser Schritt ist mit seinem Wechsel zwischen Parameteränderung und Bewertung hochgradig iterativ. In der vorliegenden Arbeit kommt hinzu, dass je Parametereinstellung Klassifikationsläufe genügend oft wiederholt werden (hier 100 Mal). Dabei wird die Auswahl von Samples für das Training und den Test jeweils zufällig, aber unter Erhalt der Klassenverteilung (stratifizierte Zufallsstichprobe) getroffen. Somit ist ausgeschlossen, dass das Vorhersageergebnis aus einer günstigen oder ungünstigen Sampleselektion resultiert. In jedem Lauf beträgt der Anteil der Trainingsdaten 70 Prozent, der der Testdaten 30. (Auch andere Aufteilungen sind denkbar, solange ausreichend viele Messungen bzw. Störungen in jeder Menge enthalten sind.)

Zur Teilung zwischen Trainings- und Testmenge werden bei allen Weichen Samples aus dem gesamten Zeitraum gezogen. Dies entspricht zwar noch keiner Prognose unter realen Bedingungen, weil dabei eben ein bekannter Zeitraum aus der Vergangenheit zur Vorhersage der unbekannten Zukunft genutzt werden würde. Es ist aber zunächst wichtiger, möglichst viele Phänomene des Abnutzungsverhaltens und der Störungsentwicklung in die Analyse einzubinden. Denn die Rohdatenbasis enthält lediglich 28 Weichen mit nur 63 Störungen, die wiederum zahlreiche Ursachen haben können. Bei der geringen Anzahl von Referenzen ist eine breite Trainingsmenge für den Lernprozess essentiell, um überhaupt erst einmal das Potenzial des Ansatzes zu ergründen.

Um die vielen Parametereinstellungen der Verfahren übersichtlich vergleichen zu können, wird eine singuläre Metrik verwendet. Die Wahl fällt auf den MCC bzw. dessen arithmetisches Mittel aus den Wiederholungen, weil er eine besonders feine Differenzierung erlaubt.

6.2.1.1 Parametrierung für k-nächsten-Nachbarn (kNN)

Bei kNN können im Wesentlichen nur zwei Größen parametrisiert werden. Die wichtigste ist k , also die Anzahl der nächsten Nachbarn, die für die euklidische Distanzberechnung und Klassifikation herangezogen werden (vgl. Abbildung 3-2). Als zweite Option können die k Nachbarn mit ihrem Abstand zum fraglichen Datenpunkt gewichtet werden, so dass die Klassen von ähnlicheren Nachbarn bei der Klassifikationsentscheidung stärker berücksichtigt werden. Angefangen bei nur einem Nachbarn (gewichtet und ungewichtet $MCC = 0.986$) bis hin zu 20 nächstgelegenen Samples (gewichtet 0.963 und gleichwertig 0.923) zeigt das Verfahren ausgezeichnete Ergebnisse, wie Abbildung 6-2 veranschaulicht. Das Maximum liegt bei einem und zwei Nachbarn, wobei bereits ab $k = 2$ die gleichwertige Behandlung der Nachbarn etwas schlechter abschneidet. Mit zunehmendem k lässt der Erfolg geringfügig aber stetig nach, ohne Gewichtung intensiver als mit Gewichtung. Die Samples verschiedener Klassen liegen im Datenraum

offenbar so dicht beieinander, dass schon bei mehr als zwei Nachbarn die Genauigkeit abnimmt. Dies ist zum Teil eine Folge der vielen Attribute, wodurch das Distanzmaß in diesem Verfahren an Aussagekraft verliert. (Dieses Phänomen ist auch als Fluch der Dimensionen bekannt, siehe [BGRS99, HAK00, DK08].) Da nichts für eine Umkehr dieses Trends spricht, wird k nicht weiter erhöht. Der Erfolg mit zwei gewichteten Nachbarn ist ohnehin beachtlich und zudem überaus stabil, da die Standardabweichung aus den Wiederholungen hier nur 0.0007 beträgt. Insgesamt liefert kNN sehr stabile Ergebnisse. Selbst die höchste Standardabweichung beträgt nur 0.0032.

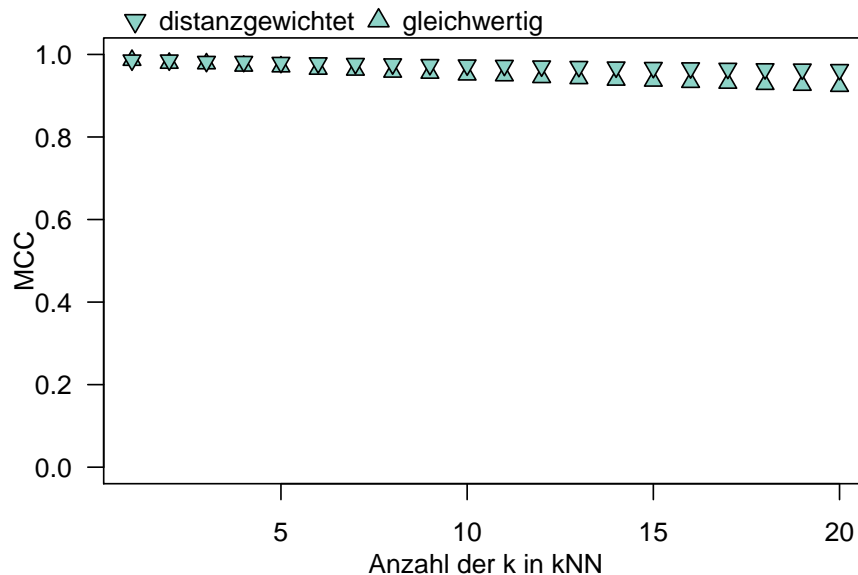


Abbildung 6-2: Prädiktionserfolg von kNN bei verschiedenen k -Werten

6.2.1.2 Parametrierung des künstlichen Neuronales Netzes (aNN)

Künstliche Neuronale Netze können auf sehr vielfältige Weise parametriert werden (vgl. Abschnitt 3.2.3). Als grundlegende Netzstruktur wird hier ein Multilayer-Feedforward-Netz mit dem Standard-Lernalgorithmus von Riedmiller und Braun [RB93] verwendet. Die Anzahl der verdeckten Schichten und der verwendeten Neuronen je Schicht sind frei wählbar. Die Suche nach den besten Netzeinstellungen erfolgt systematisch, indem die Anzahl der verdeckten Schichten schrittweise von 1 erhöht wird, während die Neuronenanzahl jeweils von 8 bis 38 inkrementiert wird. (Das Training für jede Einstellungsvariante wird nach 100 Epochen beendet, doch schon nach 60 bis 80 Epochen ändert sich die Neuronengewichtung fast nicht mehr.) Aus Abbildung 6-3[a] wird ersichtlich, dass jegliche Einstellungen mittelmäßige Ergebnisse von ähnlichem Niveau erzeugen, deutlich schlechter als zuvor bei kNN. Bei genauerer Betrachtung im Intervall des MCC-Minimums und -Maximums ($[0.256, 0.325]$) lässt sich gut erkennen, dass ab einer verdeckten Schicht die Ergebnisse nachlassen (siehe Abbildung 6-3[b]). Den Maximalwert markieren 28 Neuronen, aber die Unterschiede von 24 bis 34 sind kaum signifikant. Mit Standardabweichungen von 0.0127 bis 0.0322 ist die Performance zwar stabil, aber deutlich weniger als bei kNN. Wie schon bei der kNN-Parametrierung erwähnt, liegen die Samples unterschiedlicher Klassen im Datenraum dicht beieinander bzw. überlappen stark. Dadurch wird schnell eine bestimmte Struktur erlernt, die aber nicht global ist. Wahrscheinlich begründet sich das relativ schlechte Abschneiden dann auch mit der Tendenz zur Überanpassung und lokalen Fehlerminimierung, die für aNN typisch ist. Seit den ersten aNN in den 1940er Jahren wurden sehr viele Netztypen

entwickelt, die zum Teil sehr spezielle Formen annehmen können. Es kann daher nicht gänzlich ausgeschlossen werden, dass ein Spezialnetz existiert, das andere Eigenschaften ausweist als etablierte Typen und bessere Ergebnisse erzielen könnten. Allerdings scheint eine erschöpfende Suche angesichts der guten Ergebnisse von kNN ineffizient.

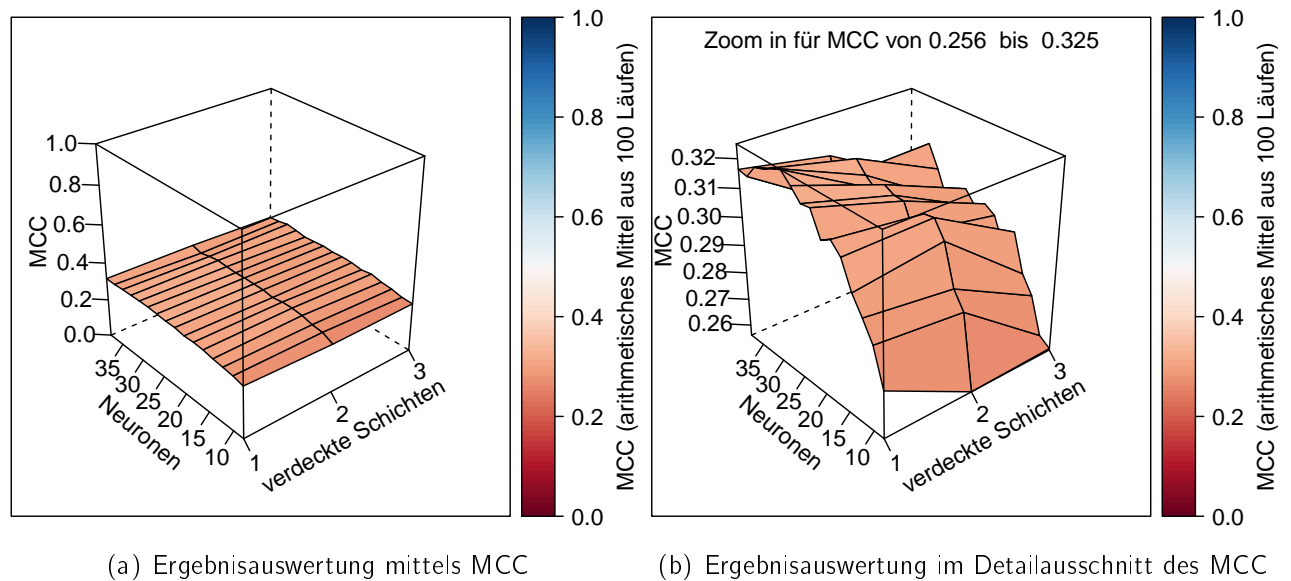


Abbildung 6-3: Prädiktionserfolg bei unterschiedlicher Anzahl von verdeckten Schichten und Neuronen im aNN

6.2.1.3 Parametrierung des Entscheidungsbaums (DTree)

Entscheidungsbäume besitzen wenige Einstellungsmöglichkeiten, lediglich das Teilungskriterium für die Verzweigung und die Option zum Beschneiden der Zweige (vgl. Abschnitt 3.2.4). Die hier verwendete Implementierung eines Entscheidungsbaums folgt dem Standardverfahren C4.5 von Quinlan [Qui93] und dessen Optimierung aus Shafer et. al [SAM96]. Um in einem Knoten die Datenmenge zu teilen, kann zwischen der Gain Ratio (unter Nutzung der Entropie) und dem Gini Index gewählt werden. In Abbildung 6-4 ist zu sehen, dass der Gini Index ($MCC = 0.940$) der Gain Ratio ($MCC = 0.868$) überlegen ist. In beiden Fällen führt ein Beschneiden zu schlechteren Ergebnissen, obwohl dies im Allgemeinen nicht so ist. Der Grund ist auch hier die starke Ähnlichkeit von Samples, so dass der Baum weit verzweigt werden muss, damit eine hohe Genauigkeit erreicht wird. Die Performance von DTree ist mit einer Standardabweichung von 0.0015 beim Gini Index und 0.0023 bei der Gain Ratio ebenfalls sehr stabil. Damit ist der Prognoseerfolg bei DTree fast so hoch wie bei kNN.

6.2.1.4 Parametrierung der Stützvektormaschine (SVM)

Bei der Stützvektormaschine muss der Anwender zunächst die Kernel-Funktion für die Transformation in den höherdimensionalen Raum wählen (siehe Abschnitt 3.2.5). Außerdem ist die Strafkonstante ($\text{Penalty} \in \mathbb{R}^{>0}$) zu parametrieren, die angibt, wie schwer ein Datenpunkt wiegt, der die Separationsmarge (vgl. Abbildung 3-5) überlappt. Hinzu kommen ganz unterschiedliche Parameter der Kernel-Funktionen, z.B. bei einem Polynom, dessen Grad, Skalar und Restkonstante oder beim Radial-Base-Function (RBF) Kernel die Weite der Gaussform ($\text{sigma } \sigma \in \mathbb{R}^{>0}$).

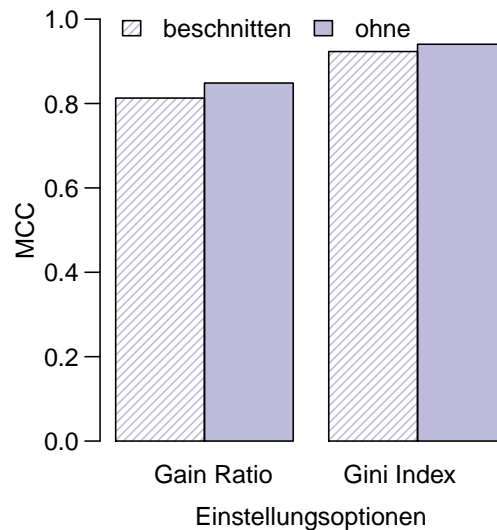


Abbildung 6-4: Prädiktionserfolg bei verschiedenen Teilungskriterien im DTree, jeweils mit und ohne Beschneidung

Auch in der hier genutzten SVM-Implementierung nach Platt [Pla99] bzw. der Verbesserung nach Keerthi et. al [KSBM01] sind diese Parameter festzulegen. Anfangs wird der lineare Kernel bei verschiedenen Werten der Penalty getestet. Weil es der einfachste Kernel ist und nur der Penalty-Wert variiert werden muss, empfiehlt es sich, mit diesem Kernel zu beginnen (vgl. [BHW10]). Bei verschiedenen Strafkosten liegt der MCC im Bereich zwischen 0.140 und 0.181 mit Standardabweichungen unterhalb von 0.01. Diese Performance ist nicht gut, was allerdings zu erwarten war, da die Klassengrenzen schließlich nicht linear verlaufen. Bessere Ergebnisse könnte der RBF Kernel erreichen. Um dessen beste Einstellungen zu finden, wird eine Rastersuche für die Parameter (Penalty und sigma) nach den Vorbildern von Hsu et. al [HCL03] sowie Min und Lee [ML05] durchgeführt. Die in Abbildung 6-5[a] dargestellte Performance verdeutlicht, wie entscheidend die Parameterwahl bei SVM ist. Von der zufälligen Klassenzuordnung ($MCC = 0$) bis zu einer sehr guten Klassifikation ist alles möglich. Die feinere Rastereinteilung offenbart, dass die maximale Performance mit einer Strafkosten von 128 und sigma von 0.2 erreicht wird (siehe Abbildung 6-5[b]). Damit erzielt der RBF Kernel nicht nur genauso hervorragende Ergebnisse wie kNN, sondern ist mit Standardabweichungen von 0.0004 im MCC-Maximum und weniger als 0.0019 im gesamten Raster auch ebenso stabil. Mit solch einem Ergebnis lohnt sich die Untersuchung eines Kernels mit polynomieller oder sigmoider Funktion nicht. Beide benötigen einen deutlich höheren Rechenaufwand und besitzen zudem mehr Parameter, erreichen jedoch oft schlechtere Ergebnisse [HCL03, S. 4], [ML05, S. 610], [BHW10, S. 4].

Insgesamt erzielen drei von vier untersuchten Verfahren gute bis ausgezeichnete Ergebnisse, wenn sie entsprechend parametrisiert werden, d.h. :

- kNN bei $k = 2$ und Gewichtung der Distanz zu diesen beiden Nachbarn,
- DTree mit Gini Index als Teilungskriterium und ohne Beschneidung,
- SVM mit dem RBF Kernel, der Strafkosten 128 und $\sigma = 0.2$.

Einzig das aNN wird den Ansprüchen an eine zuverlässige Klassifikation nicht gerecht. Deshalb wird im weiteren Verlauf auf aNN verzichtet, wenn die Ergebnisse detaillierter analysiert werden.

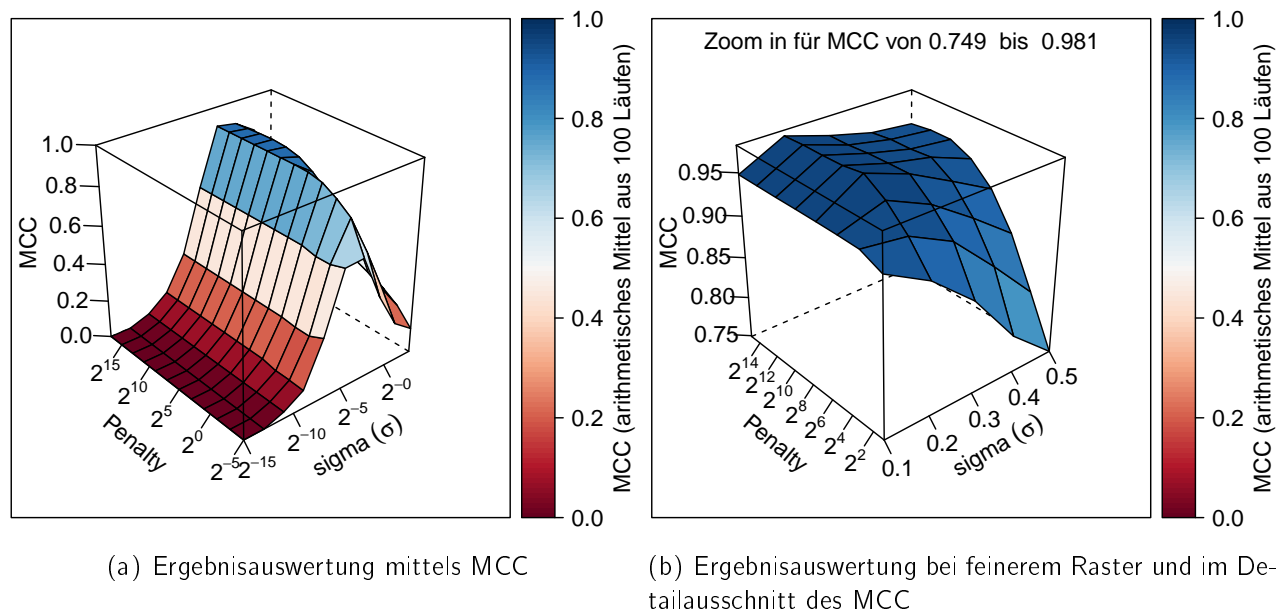


Abbildung 6-5: Prädiktionserfolg von SVM mit RBF Kernel und verschiedenen Einstellungen

An dieser Stelle sei daran erinnert, dass Lern- und Testsamples bisher datumsunabhängig aus der gesamten Rohdatenbasis gezogen wurden. Das entspricht zwar nicht einer Prognose unter realen Bedingungen. Es ist aber bei der relativ geringen Zahl an Störungsreferenzen notwendig, um beim Lernen möglichst umfangreich die Phänomene der Abnutzung und der verschiedenen Störungsursachen zu berücksichtigen und die Genauigkeit optimieren zu können. Unter Umständen werden unter realen Bedingungen weniger genaue Vorhersagen erreicht, bspw. weil eine Störung prognostiziert werden muss, deren Klassifikationsmuster nicht in den gelernten Daten enthalten war. Dies oder Ähnliches wird im späteren Verlauf des Kapitels noch analysiert.

6.2.2 Vergleich des Prognoseerfolgs bei bestmöglicher Parametrierung der Verfahren

Nach der individuellen Maximierung des Prognoseerfolgs sollen nun die Ergebnisse im Detail miteinander verglichen werden. Hierbei steht die Frage im Mittelpunkt, wie gut die einzelnen RUL-Klassen bestimmt werden konnten. Zu diesem Zweck wird die Konfusionsmatrix betrachtet, welche hier in den Abbildungen 6-6, 6-7 und 6-8 visualisiert ist. (Die vollständigen Matrizen finden sich in Anhang C.1 in den Tabellen C-1, C-2 und C-3). Dargestellt ist der Anteil jeder Zelle an der tatsächlichen RUL-Klasse. Entlang der Diagonalen offenbart sich so auch gleich die TP Rate, also der Anteil mit dem die Klasse richtig prognostiziert wurde. Neben der schlechteren Performance des Entscheidungsbaumverfahrens, die schon im vorherigen Abschnitt sichtbar wurde, sind folgende Aspekte auffällig:

- Es gibt keine wesentlichen Fehlklassifikationen bei weit auseinanderliegenden Restnutzungsdauern, da abseits der Diagonalen nur sehr kleine Werte auftreten.
- Gerade bei den kleinen Klassen z_1 bis z_4 mit einer RUL unterhalb von 0.5 Tagen ist die Prognose weniger treffsicher. Die entsprechende Missklassifikation verteilt sich aber auf dieselben Klassen. Die Ursache liegt zum einen im engeren Schwankungsbereich von Messwerten in den kleinen Klassen, so dass Ausreißer eher in eine andere Klasse sortiert werden.

Zum anderen enthalten die kleinen Klassen weniger Samples (siehe Tabelle 6-1). Dadurch ist das Lernen schwieriger, insbesondere für DTree. Hier hat kNN einen Vorteil, weil es eben nur die zwei ähnlichsten Punkte verwendet, anstatt einer allgemeinen Regel. Dafür ist SVM bei allen anderen Klassen genauer, wie die in Tabelle 6-2 zusammengefassten TP Raten unterstreichen.

Es leitet sich die Schlussfolgerung ab, dass die Prognose verlässlicher werden könnte, wenn die kleinen Klassen zusammengefasst werden würden.

$TP_r(I)$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
kNN	.768	.726	.79	.879	.93	.952	.97	.983	.986	.991	.991	.991	.992	.998
DTree	.056	.160	.3	.417	.567	.701	.80	.842	.867	.885	.875	.904	.923	.972
SVM	.724	.623	.73	.858	.946	.95	.98	.986	.991	.994	.994	.995	.995	.999

Tabelle 6-2: TP Rate je Klasse im Vergleich der Verfahren (Die 0 vor Dezimalstellen wurde weggelassen.)

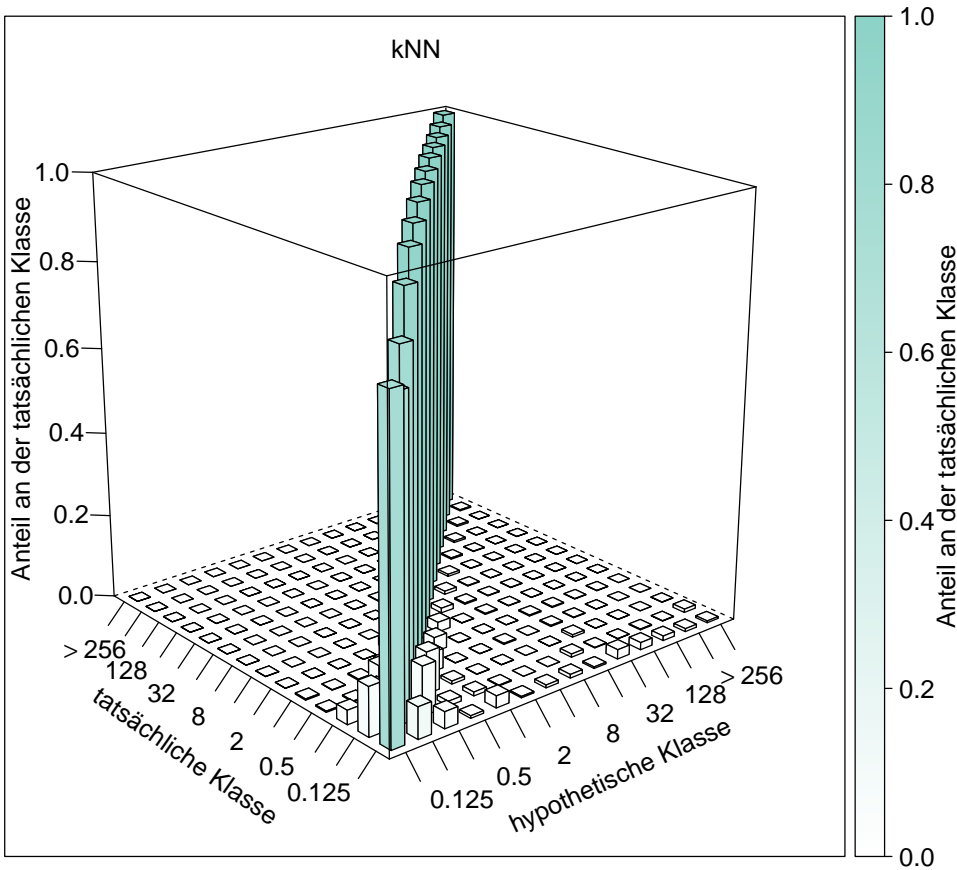


Abbildung 6-6: Konfusionsmatrix der Prognose mit kNN als Anteil an der tatsächlichen RUL-Klasse

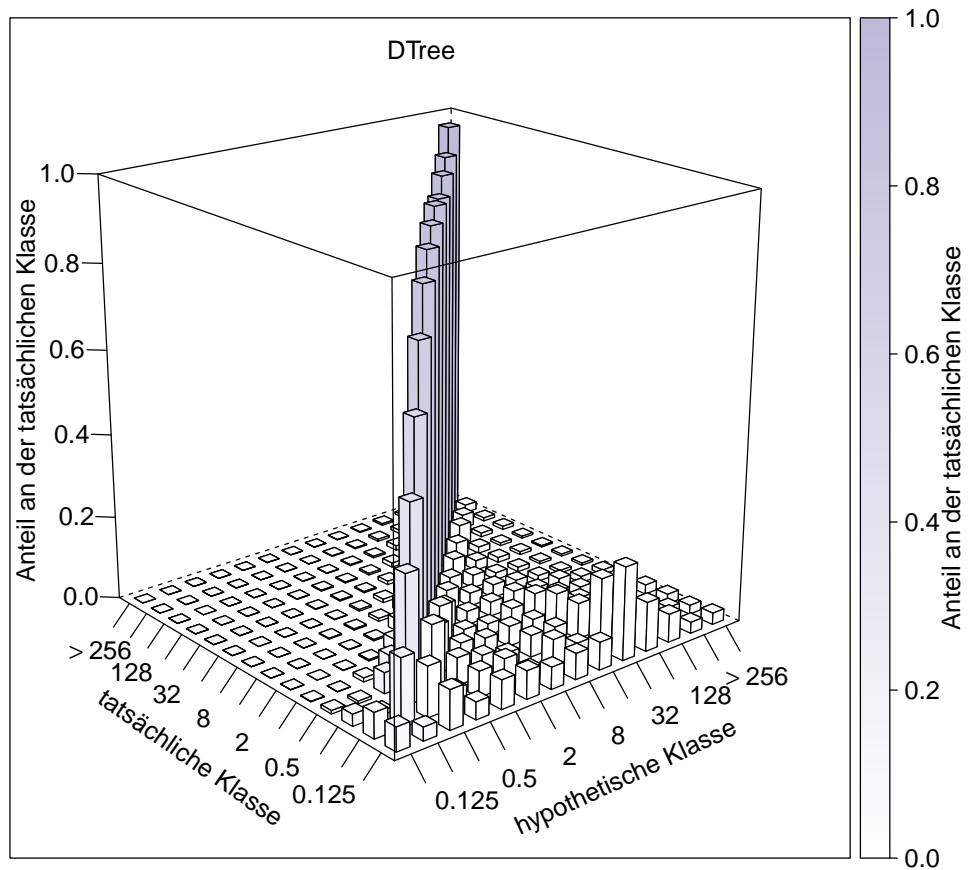


Abbildung 6-7: Konfusionsmatrix der Prognose mit DTree als Anteil an der tatsächlichen RUL-Klasse

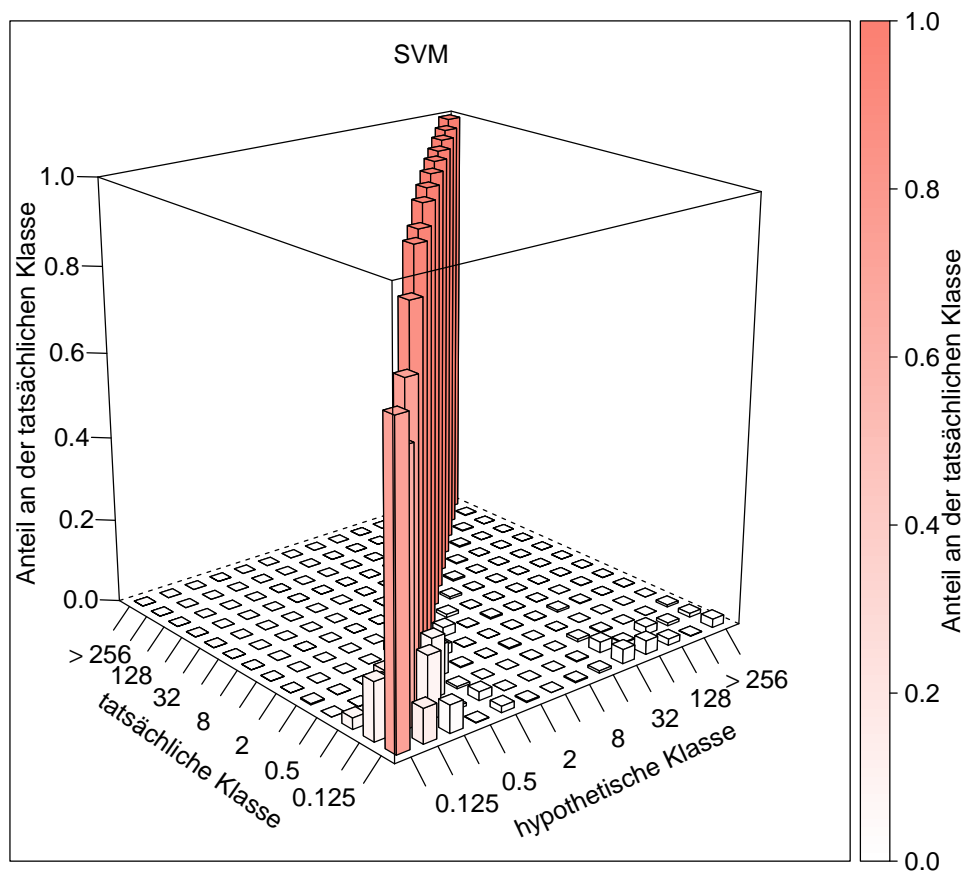


Abbildung 6-8: Konfusionsmatrix der Prognose mit SVM als Anteil an der tatsächlichen RUL-Klasse

6.2.3 Effekte alternativer RUL-Klassen auf den Prognoseerfolg

Die detaillierte Analyse im vorangegangenen Abschnitt hat ergeben, dass die Prognose einer RUL unterhalb eines halben Tages weniger zuverlässig ist. Dies gefährdet die Nutzerakzeptanz. In diesem Abschnitt soll daher untersucht werden, wie sich das Ergebnis verändert, wenn alternative Klassen gebildet werden. Das betrifft zwar in erster Linie die kleinen Klassen z_1 bis z_4 . Es ist aber auch interessant zu untersuchen, wie sich stark veränderte Klassenstrukturen auswirken.

Die RUL-Klassen wurden bisher so definiert, dass sie eine mathematische Folge bilden und weit entfernte Störungen in größeren Klassen zusammenfassen. Zudem berücksichtigen sie die Vorlaufzeiten, welche nach Aussage von Instandhaltern für die Notfallentstörung und die regulär geplante Maßnahme benötigt werden (vgl. Abschnitt 4.2.1 bzw. [Böh13c]). Diese Bedingungen werden nun aufgeweicht. Es wird aus Platzgründen darauf verzichtet, die Definitionsgleichungen explizit aufzuführen. Stattdessen werden die Intervallstufen genannt. Die Gleichungen finden sich jedoch im Anhang C.2 und sind in ihrer Form analog zur Gleichung (6.1), welche die ursprüngliche Struktur definierte (folgend als Z_0 referenziert).

- Die drei kleinsten Klassen werden zusammengefasst, so dass unterhalb 0.25 Tagen nicht weiter differenziert wird. Die Struktur wird als Z_1 beschrieben.

$$Z_1 \quad \text{mit } \{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, \infty\} \quad (6.11)$$

- Es werden Klassen manuell definiert, die sich stärker am Kalender orientieren. Die kleinste Klasse ist von drei Stunden begrenzt und trägt damit der Notfallentstörung Rechnung. Ein Tag soll Zeit geben innerhalb von 24 Stunden reagieren zu können. Drei und fünf Tage folgen dem Feinplanungszeitraum, während 14 und 28 Tage vollständige Wochen repräsentieren. Danach werden rund zwei, vier und acht Monate als Intervalle festgelegt. Diese Aufteilung kommt den Bedürfnissen der Weicheninstandhaltung am nächsten.

$$Z_2 \quad \text{mit } \{0.125, 1, 3, 5, 14, 28, 61, 122, 244, \infty\} \quad (6.12)$$

- Die Klassen unterhalb von zwei Tagen und oberhalb von 64 Tagen werden jeweils zusammengefasst. So entstehen bei Z_3 wenige Klassen.

$$Z_3 \quad \text{mit } \{2, 4, 8, 16, 32, 64, \infty\} \quad (6.13)$$

- Sehr wenige Klassen zu bilden, steht auch bei der Struktur Z_4 im Mittelpunkt, hier allerdings an den Zeiten für Notfallentstörung und Feinplanung ausgerichtet.

$$Z_4 \quad \text{mit } \left\{\frac{2}{24}, 3, 6, 60, \infty\right\} \quad (6.14)$$

- Statt einer geometrischen Folge wird der Klassenstruktur Z_5 eine arithmetische Folge zugrundegelegt. Intervalle werden in äquidistanten Abständen von fünf Tagen gesetzt.

$$Z_5 \quad \text{mit } \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, \infty\} \quad (6.15)$$

- Als letzte Struktur werden eine geometrische und eine arithmetische Folge kombiniert.

Von 3 Stunden bis 8 Tage werden die Zweierpotenzen für die Intervalle genutzt, danach ein gleichmäßiger Abstand von je acht Tagen.

$$Z_6 \quad \text{mit } \{0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, \infty\} \quad (6.16)$$

Für die Performanceauswertung wird die ähnlichkeitsgewichtete, makroskopische AUC verwendet. Sie erlaubt als singuläre Metrik einen schnellen, übersichtlichen Vergleich, der zudem invariant gegenüber der Klassenverteilung ist. Das ist wichtig, weil mit der unterschiedlichen Klassenanzahl auch die Samples je Klasse variieren. Außerdem betont die $wAUC$ eine gute Unterscheidung von per se ähnlichen Klassen. So wird besonders Wert darauf gelegt, die reduzierte Missklassifikation zwischen den kleinen RUL-Klassen anzuzeigen.

Die Prognoseergebnisse, die die einzelnen Verfahren bei der jeweiligen Klassenstruktur erzielen, sind in der Tabelle 6-3 zusammengefasst. SVM erzielt bei den Strukturen Z_4 und Z_6 die schlechtesten Ergebnisse. Allerdings ist dort keines der Verfahren erfolgreich. Zum einen verwendet Z_6 zu viele Klassen mit zu kleinen RUL Intervallen, als dass sich diese gut über Messungen abbilden lassen. Die Messungen überlappen zu sehr die Klassengrenzen. Zum anderen bildet Z_4 zu wenige Klassen, die noch dazu mit 6 und 60 eine schlechte Einteilung treffen. Diese beiden größten Klassen decken nahezu alle Messbereiche ab. Dadurch kann kein Verfahren eine gute Zuordnung erreichen. In Z_5 gibt es keine Klassen mit kleiner RUL, weshalb sich der Erfolg wieder verbessert. Jedoch verhindern die hohe Klassenanzahl und die Äquidistanz exzellente Ergebnisse. Z_3 erzielt zwar noch sehr gute Ergebnisse, aber sieben Klassen scheinen bereits zu wenig zu sein. Den höchsten Erfolg erreicht Z_2 . Dass eine Struktur so gut abschneidet, die keiner bestimmten mathematischen Form folgt, ist nicht verwunderlich. Denn die RUL hängt nicht von der Zeit selbst ab, sondern von den Verschleißbedingungen.

Struktur	Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Klassenanzahl	14	12	10	7	5	18	18
$wAUC$ kNN	0.99152	0.99516	0.99651	0.96496	0.22079	0.88583	0.30927
$wAUC$ DTree	0.92254	0.98641	0.95298	0.95342	0.20797	0.62360	0.28165
$wAUC$ SVM	0.98963	0.99468	0.99589	0.96712	0.12727	0.84086	0.22440

Tabelle 6-3: Vergleich der ähnlichkeitsgewichteten, makroskopischen AUC je Klassenstruktur und Verfahren

Es ließen sich sicher noch zahlreiche weitere Klassenstrukturen erdenken und testen, aber dies ist hier nicht sinnvoll. Denn letztlich muss der Anwender, also der Instandhalter bestimmen, welche Klassenstruktur für ihn gut handhabbar ist. Weil die Klassenstruktur die Prognosegüte mitbestimmt, sind vielmehr die identifizierten Anhaltspunkte wichtig, da sie eine Orientierung vorgeben. So sind ungefähr zehn Klassen eine gute Wahl, wenn die Granularität passend gewählt wird. D.h. große RUL oberhalb von 14 bis 28 Tagen sind in Klassen mit weitem Intervall zusammenzufassen. Mittlere Klassengrößen sind im Bereich von 3 bis 28 Tagen zu wählen. Für die Zeitpunkte kurz vor der Störung sollten nur sehr wenige Klassen gebildet werden, die im Intervall von 12 bis 24 Stunden liegen. Dadurch ist zwar die RUL-Auflösung geringer, aber die Prognose zuverlässiger. Fortan wird in der vorliegenden Arbeit deshalb auch die Struktur Z_2 verwendet. Die Anforderung nach einer Zustandsprognose in Form einer RUL (siehe Nummer 1 in Tabelle 5-9) ist damit allemal erfüllt, so dass Instandhaltungsmaßnahmen gut geplant werden

können. Die Reduktion der Klassen hat, neben dem leicht verbesserten Ergebnis, noch einen weiteren Vorteil. Die Trainings- und Testphase für DTree und SVM verkürzen sich.

6.2.4 Performance Evaluation der Prognose unter ökonomischen Gesichtspunkten

Im Anschluss an den Performancevergleich der Prognose bei bestmöglicher Parametrierung und bei verschiedenen Klassenstrukturen folgt nun die Analyse unter ökonomischen Gesichtspunkten. Analog zur Evaluation von SIDIS W im Kapitel 5.1.5 wird die Break Even Situation mit der in Abschnitt 4.2 entwickelten Methode untersucht und visualisiert. Dadurch soll nachgewiesen werden, dass die Vorhersage auch zur Senkung von Instandhaltungskosten beiträgt. Die Größen für die Fehlalarmkosten und den Nutzen einer rechtzeitig erkannten Störung, welche bereits zur Evaluation des Diagnosesystems ermittelt wurden (siehe Tabelle 5-4), sind dabei nach wie vor gültig. Und wie zuvor wird mit einem schlechtesten und einem bestmöglichen Verhältnis zwischen Kosten und Nutzen (Kosten-Nutzen-Verhältnis (engl. Cost-Benefit-Ratio) (CBR)) gearbeitet.

Zunächst sind die zehn RUL-Klassen in binäre Klassen umzuwandeln. Klassen mit einer RUL von fünf oder weniger Tagen werden dazu als positive Klasse zusammengefasst (vgl. Abschnitt 4.2.1). Alle restlichen Klassen werden als negative Klasse subsumiert. Wird eine Messung zur positiven Klasse zugeordnet, so sollte eine Instandhaltungsmaßnahme geplant und in den kommenden fünf oder weniger Tagen durchgeführt werden, um die Störung zu verhindern. Bei binären Klassen werden nun wieder die originäre FP Rate und die TP Rate verwendet, anstelle der $macroFP_r$ und der $macroTP_r$. Letztere wurden für die konsistente ROC-Metrik bei multiplen Klassen benötigt. Bei binären Klassen erreicht kNN eine TP Rate von 0.987, DTree erzielt 0.826 und SVM kommt auf 0.995. Alle drei Verfahren erfüllen somit die Anforderung (Nummer 2 in Tabelle 5-9) nach einer TP Rate von mindestens 0.5 in der Klasse mit einer RUL von fünf oder weniger Tagen. Wie bereits im Abschnitt 5.1.5 festgestellt wurde, geht aus den verfügbaren Daten nicht hervor, wie viele Verspätungsminuten (VMin) jede Störung verursacht hat. Daher wird das arithmetische Mittel der gesamten VMin aller 29 Weichen aus den Jahren 2010 bis 2013 verwendet. Die DB Netz AG ermittelte hierfür 47.3 VMin je Störung. Bei TP Raten über 80 Prozent kann zu Recht angenommen werden, dass mit der Vorhersage auch über 50 Prozent der VMin verhindert werden können. Dies gilt selbst dann, wenn anstatt der 63 Störungen aus der selektierten Rohdatenbasis alle 98 Störungen zugrunde gelegt werden ($63 \cdot 0.8 \approx 50 > 49 = 98 \cdot 0.5$). Somit wird auch diese Anforderung (Nummer 3 in Tabelle 5-9) erfüllt.

Als nächstes wird die Break Even Situation auf Basis des ROC-Graphs visualisiert. Im Unterschied zu SIDIS W müssen aber nicht mehrere ROC-Punkte nach bestimmten Maßstäben erzeugt und ihre konvexe Hülle gebildet werden. Die Linie bzw. ROC-Kurve ergibt sich direkt aus den Prognoseverfahren und wird unmittelbar in die dritte Dimension des Störungsanteils (T_r) erweitert, um eine Fläche zu erhalten. Aus den Verfahren resultiert zu jeder Messung ein Wert für die jeweilige Klassenzugehörigkeit. Mit diesem Wert wird die ROC-Kurve regulär gebildet, indem unterschiedliche Niveaus angenommen werden, ab welchen eine Messung der positiven oder negativen Klasse zugeordnet wird (vgl. Abschnitt 3.3.1 oder Abbildung A-1 in Anhang A.2).

In der Abbildung 6-9 werden die Break Even Situationen für jedes Verfahren und zum Vergleich für SIDIS W dargestellt. Die Auflösung von $T_r \leq 0.001$ wurde gewählt, da Störungen nur sehr sel-

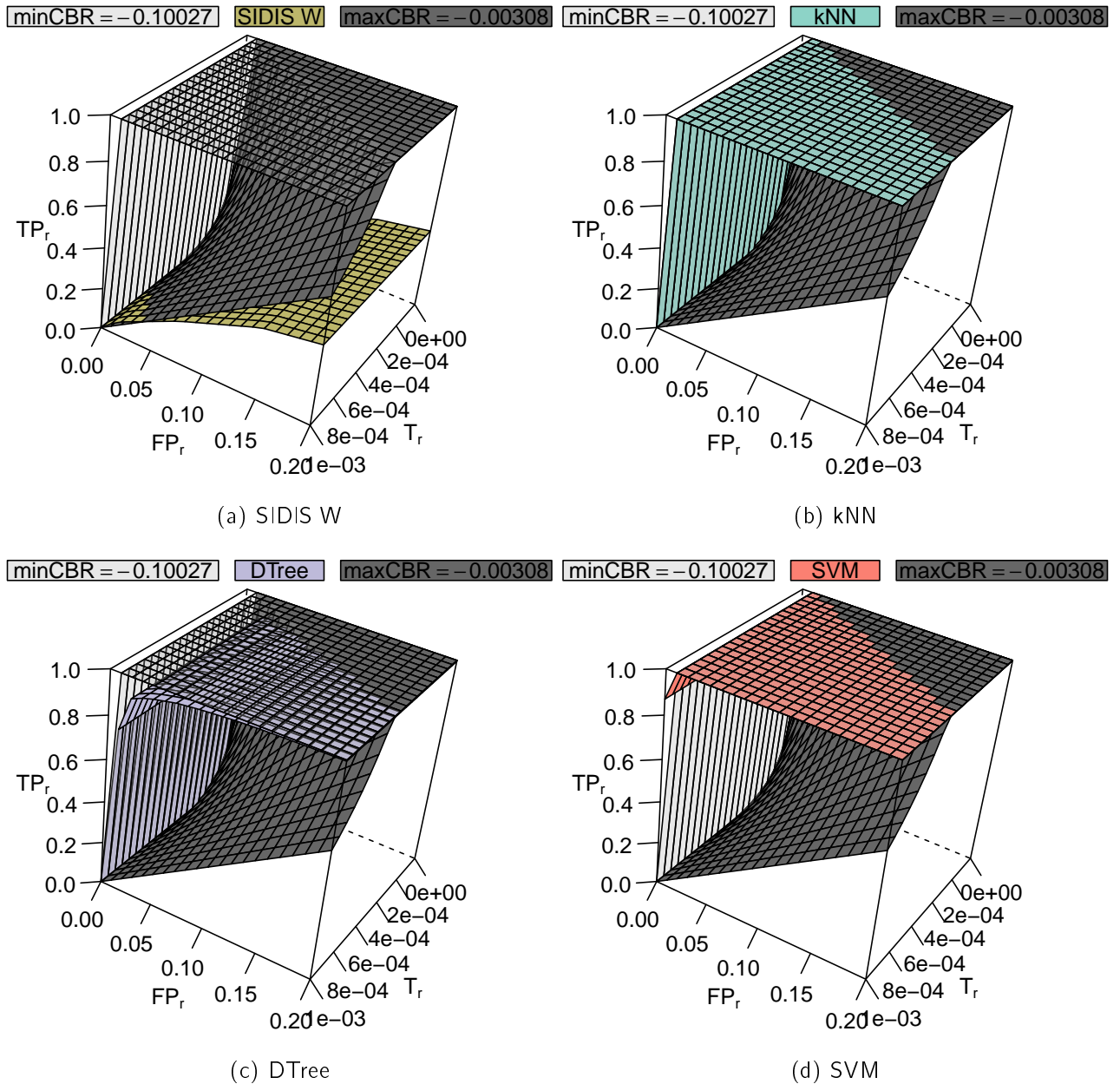


Abbildung 6-9: Visualisierung der Break Even Situation der Störungsprognose, Auflösung im Intervall FP Rate bis 0.2 und wenn bei jedem 1000. Umlauf oder seltener eine Störung eintritt (T_r); bei positivem Nutzen liegen die farbigen Flächen links bzw. oberhalb der anderen Flächen

ten mehr als ein Mal alle 1000 Umläufe vorkommen, wie bereits hinlänglich im Abschnitt 5.1.5 erläutert wurde. (Für eine bessere Sichtbarkeit der Flächen ist außerdem die FP_r im nur im Intervall von 0 bis 0.2 abgebildet.) Alle drei Prädiktionsverfahren können Instandhaltungskosten senken, wenn die Kosten eines Fehlalarms minimal und der Satz je VMin möglichst hoch sind ($maxCBB$) sowie Störungen nicht zu selten auftreten. Ihre ROC-Flächen liegen größtenteils links oberhalb der $maxCBB$ -Fläche. Allerdings erreicht DTree schon nicht mehr die Wirtschaftlichkeit, wenn das schlechteste Verhältnis aus Kosten und Nutzen ($minCBB$) herangezogen wird. Seine ROC-Fläche verläuft zwischen der transparenten $minCBB$ -Fläche und der $maxCBB$ -Fläche. Nur kNN und SVM liefern eine ausreichend genaue Prognose, um auch wirtschaftlich zu funktionieren, unter Umständen bis hin zu circa einer Störung alle 13000 Umläufe ($T_r = 0.000075$). Die Anwendung dieser beiden Prognoseverfahren ist wirtschaftlich und würde dem Infrastrukturbetreiber mehr Verspätungskosten einsparen, als Kosten durch unnötige Wartungsmaßnahmen aus Fehlalarmen entstünden. Somit erfüllen k-nächste-Nachbarn und die Stützvektormaschine die entsprechende Anforderung (Nummer 4 in Tabelle 5-9).

Mit der gewählten Klassenstruktur und den optimalen Parametereinstellungen zeigen die drei Verfahren ein erhebliches Potenzial für eine zuverlässige und präzise Zustandsprognose. Die Vorhersage mittels kNN und SVM arbeitet zudem wirtschaftlich. Alle an die Prädiktion gestellten Anforderungen werden theoretisch erfüllt. Offen bleibt die Frage, wie gut eine Prognose unter realen Bedingungen ist. Sie wird im kommenden Abschnitt untersucht.

6.3 Anwendung der Verfahren unter realen Bedingungen

Für das Lernen und die Evaluation der Verfahren zur datengetriebenen Zustandsvorhersage wurde bisher die gesamte Rohdatenbasis in Test- und Trainingsmenge zerlegt, unabhängig von der zeitlichen Reihenfolge. Vor dem Hintergrund einer kleinen Menge an Zustandsreferenzen war dies notwendig, um möglichst viele Phänomene der Störungsentwicklung zu berücksichtigen. Die grundsätzliche Machbarkeit einer RUL-Prädiktion konnte so nachgewiesen werden. Außerdem wurde ersichtlich, wie gut eine Vorhersage bei bestmöglicher Parametrierung sein kann und welchen Einfluss die Klassenstruktur auf das Prognoseergebnis nimmt. Doch erst die Vorhersage unter realen Bedingungen kann beweisen, ob das Konzept schon einsatzreif ist. Hierzu wird nun die Rohdatenbasis chronologisch geteilt. Messdaten vor einem bestimmten Zeitpunkt werden für das Training der Verfahren verwendet, Daten nach dem Zeitpunkt für den Prognosetest. Das Ergebnis wird mit dem MCC evaluiert, weil er auch kleine Unterschiede im Prognoseresultat darstellt, aber vor allem weil er einfach zu interpretieren ist (vgl. Abschnitt 3.3.2). So ist etwa ein MCC-Wert von 0.0 gleichbedeutend mit einer zufälligen Vorhersage, während 1 eine perfekte Klassenzuordnung bzw. -1 eine vollständig konträre Zuordnung abbilden.

Da für die einzelnen Weichen Aufzeichnungen aus unterschiedlichen Zeiträumen vorhanden sind (siehe Abbildung 5-1) und auch die Störungen unregelmäßig auf Weichen und auf Zeiträume verteilt sind, werden mehrere chronologische Teilungen vorgenommen. Das erste Datum der Teilung wurde so gewählt, dass mindestens ein Jahr seit Aufzeichnungsbeginn in die Trainingsmenge eingeht. Aus der verbleibenden Prognosemenge wurden sukzessive 66 Tage (ein Sechstel des verbleibenden Aufzeichnungszeitraums) in die Trainingsmenge verschoben. Stellstrommuster sind nicht weichenindividuell, sondern bei gleichen Bauarteigenschaften ähnlich (vgl. Abschnitt 5.2.3). Existieren für eine Weiche bisher keine Störungen in den Trainingsdaten, müssen

diese daher nicht erst gesammelt werden, sondern sie wird gemäß ihrer Eigenschaften (Weichentyp, Verschlussart, Anzahl der Zungenrollpaare) einer Gruppe zugeordnet. Wichtig ist, dass die Trainingsmenge je Bauartcluster möglichst viele Referenzen je Störungsursache enthält.

Die Tabelle 6-4 listet die Prognoseresultate auf. Für jedes Datum der Zerlegung sind darin auch die entsprechende Anzahl der gelernten und prognostizierten Tage, der Messungen und der Störungen aufgeführt. Das Ergebnis der Prognose unter realen Bedingungen ist im Vergleich mit dem bisherigen Erfolg überraschend. Unabhängig vom Datum, von dem an die Prognose vorgenommen wird, ist die Vorhersage kaum besser als eine zufällige Zuordnung zu den RUL-Klassen. (Es sei erwähnt, dass auch eine Änderung der Parametereinstellungen das Ergebnis nicht verbessert.) Wie ist dies zu begründen?

Prognose ab		Erlernt	Prognostiziert	Verfahren	MCC wenn ein Anteil der Prognosemenge zum Lernen beigemischt wird			
					0 % (Realprognose)	1 %	5 %	10 %
01.Mrz.08	Tage	363	378	kNN	0.046	0.631	0.838	0.907
	Messungen	61542	122736	DTree	-0.011	0.318	0.569	0.666
	Störungen	10	53	SVM	0.013	0.606	0.791	0.885
06.Mai.08	Tage	429	312	kNN	0.062	0.603	0.829	0.893
	Messungen	80689	103589	DTree	0.004	0.281	0.531	0.637
	Störungen	14	49	SVM	0.060	0.618	0.834	0.909
11.Jul.08	Tage	495	246	kNN	0.068	0.581	0.819	0.894
	Messungen	102297	81981	DTree	0.053	0.196	0.463	0.607
	Störungen	20	43	SVM	0.066	0.553	0.819	0.879
15.Sep.08	Tage	561	180	kNN	0.061	0.612	0.837	0.908
	Messungen	124756	59522	DTree	0.122	0.204	0.463	0.581
	Störungen	29	34	SVM	0.107	0.658	0.886	0.958
20.Nov.08	Tage	627	114	kNN	0.042	0.547	0.825	0.887
	Messungen	152094	32184	DTree	0.000	0.126	0.378	0.510
	Störungen	34	29	SVM	0.050	0.543	0.853	0.906
25.Jan.09	Tage	693	48	kNN	-0.100	0.498	0.789	0.879
	Messungen	179424	4854	DTree	0.078	0.100	0.312	0.520
	Störungen	58	5	SVM	-0.071	0.411	0.805	0.909

Tabelle 6-4: Vergleich der Prognoseresultate je Verfahren, wenn Daten vor einem bestimmten Zeitpunkt für die Vorhersage ab diesem Datum genutzt werden

Der leicht erhobene Verdacht, dass eine Überanpassung der bisherigen Vorhersage anhand der Gesamtdaten der Grund ist, kann mühelos entkräftet werden. Allein wenn beim Lernen nur ein Prozent der zu prognostizierenden Messungen beigemischt wird, verbessert sich das Ergebnis dramatisch, wie ebenfalls in Tabelle 6-4 zu lesen ist. Bei zehn Prozent erreicht die Genauigkeit schon fast wieder sehr gute Werte. Wenn mit einem so geringen Anteil von einem Prozent eine akzeptable Genauigkeit erreicht wird, ist Overfitting ausgeschlossen.

Viel wahrscheinlicher ist, dass für die Entwicklung je Störung und Bauartcluster einfach nicht genügend Referenzen vorhanden sind, damit das Konzept der zuverlässigen und präzisen Vorhersage unter realen Bedingungen einsatzbereit ist. Wie der Tabelle 6-5 entnommen werden kann, existieren nur wenige Fälle, in denen sich in einem Bauartcluster Störungen mit derselben Ursache an derselben Bauteilgruppe wiederholen. (Würden zusätzlich die Freitexte der

SAP-Störungsmeldungen interpretiert und hinzugenommen, fiel die Zahl der wiederholbaren Störungen noch viel geringer aus.) Wenn doch mehrere Störungen aufgetreten sind, lässt der Text wenig Schlüsse auf die tatsächliche Ursache zu, so dass die Referenz nicht zwingend das gleiche Störungsverhalten beschreibt.

Bauart-cluster	enthaltene Weichen	Anzahl Störungen	ProblCodeTxt (Ursache)	ProblGrpText (Bauteilgruppe)
2	14, 11, 12, 9, 13, 15, 8	14	keine Endlage	Schadbild LST Weichen
		5	nicht codiert Eintrag im Textfeld	Schadbild LST Weichen
		1	nicht stellbar	Schadbild LST Weichen
		2	reagiert nicht	Schadbild allgemein
		2	schwergängig	Schadbild allgemein
3	1, 21, 2	1	Enden verbogen	Schadbild Fahrbahn/Oberbau
		1	Stellkraft zu niedrig	Schadbild LST Weichen
		2	keine Endlage	Schadbild LST Weichen
		1	nicht codiert Eintrag im Textfeld	Schadbild Elektrotech.
		4	nicht codiert Eintrag im Textfeld	Schadbild LST Weichen
		2	nicht stellbar	Schadbild LST Weichen
		5	unzeitige Auffahrmeldung	Schadbild LST Weichen
4	7, 10, 18, 16, 5, 4, 17, 3, 19, 25, 20, 26	1	?	?
		1	Fehlfunktion	Schadbild allgemein
		3	Stellkraft zu niedrig	Schadbild LST Weichen
		4	keine Endlage	Schadbild LST Weichen
		5	nicht codiert Eintrag im Textfeld	Schadbild LST Weichen
		1	reagiert nicht	Schadbild allgemein
5	29, 27, 23, 28	1	aufgefahren	Schadbild LST Weichen
6	22	2	keine Endlage	Schadbild LST Weichen
		1	nicht codiert Eintrag im Textfeld	Schadbild LST Weichen
		4	unzeitige Auffahrmeldung	Schadbild LST Weichen

Tabelle 6-5: Anzahl der nach Ursache und Bauteilgruppe unikaten Störungen (Referenzen) je Bauart-cluster, wenn nur die definierten Standardtextfelder der SAP-Meldungen betrachtet werden

Es kommen auch noch weitere Gründe für die mittelmäßigen MCC-Werte infrage. Zum einen unterliegen die Weichen einer fortwährenden, nicht umkehrbaren Alterung und Abnutzung, wie im Abschnitt 5.2 ausführlich erläutert wurde. Für die Zustandsverschlechterungen aufgrund der Liegedauer, Verwitterung und der betrieblichen Belastung (Achsen und Achslast bei Überfahrten) existieren keine Daten. Ihr Einfluss muss unquantifiziert bleiben. Zum anderen gilt Ähnliches auch für reguläre Wartungsmaßnahmen an den Weichen, die den Zustand oder Bauteiljustierungen verändern, bspw. am Verschluss oder am Motor. Dadurch ändern sich auch Messwerte, ohne dass die Prognose unter Realbedingungen dies berücksichtigen kann.

Es lassen sich somit die folgenden Gründe für die Performance festhalten:

1. zu geringe Anzahl an Referenzen je Störungsursache und Bauartcluster bzw. Weiche,
2. nicht berücksichtigbare Langzeiteinflüsse der Alterung und Abnutzung sowie
3. nicht bewertbare Effekte unbekannter Wartungsmaßnahmen.

Dass die Performance dadurch beeinflusst wird und vor allem wie stark, lässt sich ohne deutlich umfangreichere Datenbasis nicht beweisen. Allerdings kann eine Detailbetrachtung des Prädiktionsergebnisses unter realen Bedingungen diese Aussagen untermauern. Denn es gibt durchaus einzelne Störungen die relativ gut vorhergesagt werden konnten, wie aus dem ROC-Graph in Abbildung 6-10 ersichtlich wird. (Die korrespondierende Verteilung der verschiedenen MCC-Werte befindet sich in Abbildung C-1 im Anhang C.3.) Im ROC-Graph bezeichnet jeder Punkt die RUL-Prognose, wenn die Messwerte vor jeder einzelnen Störung anhand der Messwerte einer zeitlich davorliegenden Störung gelernt und klassifiziert werden. So entstanden $1953 \cdot \left(\frac{63-1}{2}\right) \cdot 63$ Prognosen je Verfahren. Viele davon liegen erwartungsgemäß unweit der Diagonalen, also der zufälligen Prognose. Aber Vorhersagen mit einer niedrigen FP Rate und einer hohen TP Rate gehören überwiegend zu Störungsparen, die eine sehr ähnliche Ursache aufweisen, keine vier Monate auseinanderliegen und mitunter von derselben Weiche stammen oder zumindest aus dem gleichen Bauartcluster. Wohingegen Punkte im unteren, rechten Teil des Graphen hauptsächlich Störungspaare darstellen, die eindeutig aus verschiedenen Ursachen hervorgingen, z.B. Schwer- gang im Verschluss gegenüber Blockierung durch Schnee und Eis oder verbogene Zungenenden gegenüber ungeschmierten Gleitstühlen.

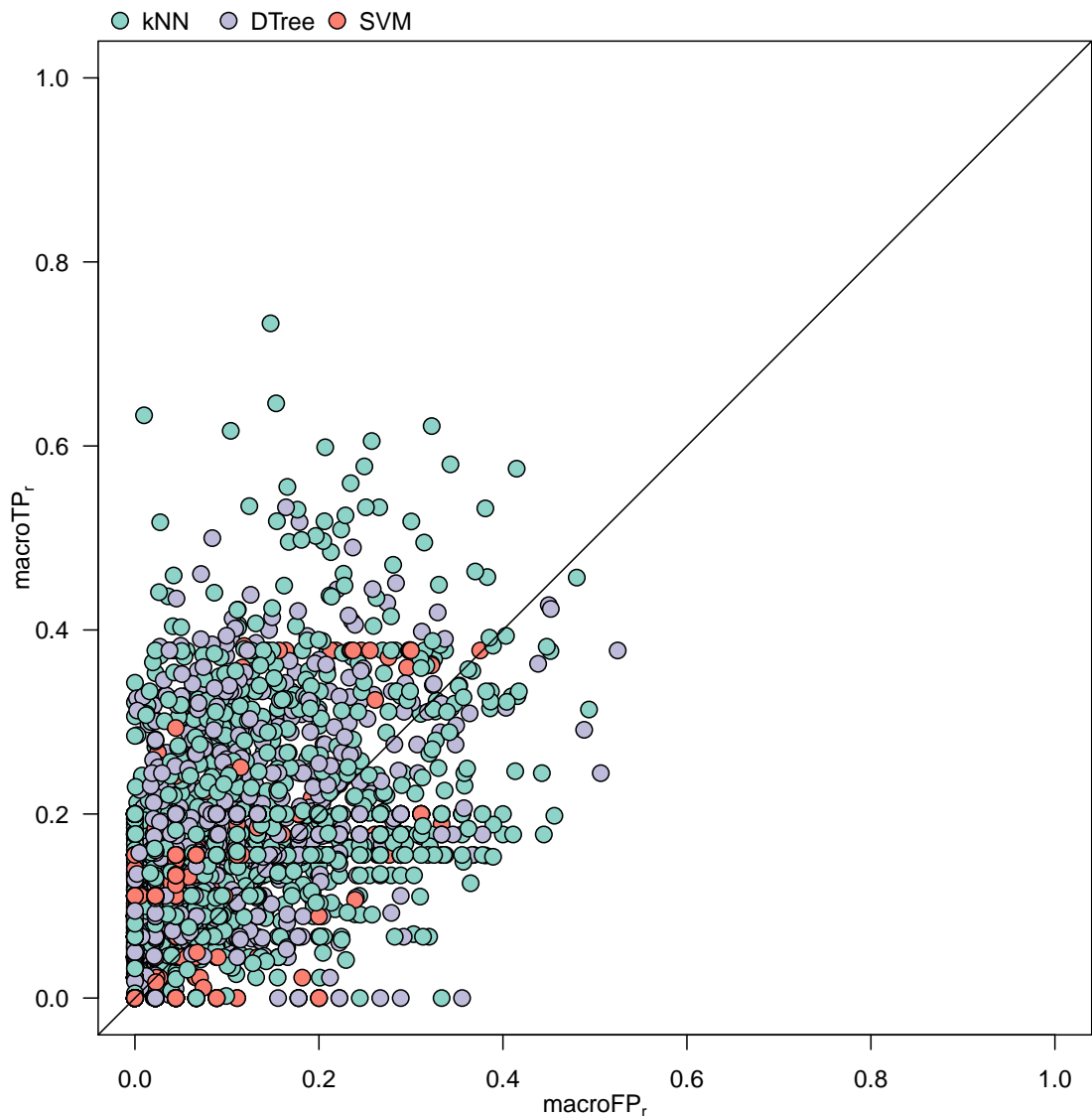


Abbildung 6-10: ROC-Graph der paarweisen Vorhersage unter realen Bedingungen von je einer Störung anhand jeder einzelnen zurückliegenden Störung

Eine systematische Analyse sollte wegen des Umfangs von rund 6000 Prognosen im Rahmen einer eigenständigen, weiterführenden Arbeit durchgeführt werden. Diese sollte sich darauf konzentrieren, die Störungsdokumentation detailliert auszuwerten. Dazu müssten jeweils acht Datenfelder manuell interpretiert werden, die zum Teil recht kryptische Freitexte enthalten. Und selbst solch eine systematische Analyse würde maximal die Begründungen stützen. Ein Beweis kann erst durch eine gelungene Prognose erbracht werden. Und dafür sind signifikant mehr Daten notwendig, so dass je Bauartcluster – oder besser je Weiche – und je Störungsursache mindestens zwei eindeutig gleiche Störungen in den Daten enthalten sind. Bestenfalls sind darüber hinaus Einflüsse von Wartung, Verwitterung und Verschleiß auszuschließen. Solch eine Datenbasis ist aber nicht verfügbar, weshalb die Einsatzreife der entwickelten präzisen und verlässlichen RUL-Prognose in zukünftigen Arbeiten nachgewiesen werden muss.

6.4 Zusammenfassung des Kapitels

Die Entwicklung und Evaluation des Prognosemodells für Weichenstörungen stand im Mittelpunkt dieses Kapitels. Dazu wurde zunächst das fundamentale Konzept genau definiert. Es wurde die stetige Restnutzungsdauer (RUL) in diskreten RUL-Klassen zusammengefasst, wodurch die Prognose in ein Klassifikationsproblem multipler Klassen transformiert werden konnte. Ein Vorteil dieses für eine RUL-Prognose ungewöhnlichen Ansatzes ist, dass für Klassifikationsaufgaben zahlreiche, leistungsfähige Verfahren aus dem Data Mining existieren. Zur Prognose neuer Messungen wird dabei die Klassenzugehörigkeit anhand der Attribute der Rohdatenbasis vorgenommen, wobei die Zuordnungsmuster zuvor anhand bekannter Messungen erlernt werden. Daraufhin wurden systematisch vier solche Verfahren ausgewählt, die sich prinzipiell bei vielen unbalancierten Klassen mit verrauschten, hochdimensionalen Daten eignen: k-nächste-Nachbarn (kNN), künstliche Neuronale Netze (aNN), Entscheidungsbaumverfahren (DTree) und Support Vector Machines (SVM). Da auch die Evaluation eine entscheidende Rolle spielt, wurden zur Problemstruktur passende Metriken (makroskopische Receiver Operating Characteristic (macroROC) und Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient) (MCC)) bestimmt bzw. neu definiert. Letzteres bezieht sich auf die gewichtete, makroskopische Fläche unter der ROC-Kurve (engl. Area Under the ROC Curve) (AUC). Sie wurde hier speziell definiert, um möglichst gut vergleichen zu können, inwieweit Prognosen benachbarte Klassen unterscheiden.

Im weiteren Verlauf des Kapitels wurde die Leistungsfähigkeit der vier Verfahren miteinander verglichen. Weil der Erfolg eines jeden Verfahrens stark von der Parametrierung abhängt, ging dem Vergleich eine individuelle Maximierung des Klassifikationsergebnisses anhand der Parameter voraus. Diese umfassende Suche und die dabei stark streuende Performance unterstreichen, wie wichtig es für die Anwendung ist, die richtigen Parametereinstellungen zu identifizieren. Dies für die Weichenzustandsprognose systematisch analysiert zu haben, stellt einen wesentlichen wissenschaftlichen Beitrag dieser Arbeit dar. Es stellte sich heraus, dass mit aNN die gegebene Problemstellung nicht zufriedenstellend zu lösen ist, weshalb diese nicht weiter betrachtet wurden. In der Detailbetrachtung der verbleibenden drei Verfahren wurde ersichtlich, dass zwar Klassen mit langer RUL ausgezeichnet klassifiziert werden konnten. Aber kleine Klassen mit weniger als einem halben Tag RUL konnten nicht so treffsicher prognostiziert werden. Hier zeigte insbesondere DTree deutliche Schwächen, während kNN und SVM noch relativ gute

Ergebnisse im Bereich einer richtig Positiv (engl. True Positive) (TP) Rate von 0.62 bis 0.79 lieferten. Als Schlussfolgerung daraus wurden alternative Klassenstrukturen untersucht. Dabei bestätigten sich weitere wichtige Vorteile der RUL-Diskretisierung. Solange Klassen mit langen RUL-Intervallen breit gefasst werden und kürzere Zeiten bis zur Störung immer enger, aber nicht zu granulär, bleibt die Vorhersagegenauigkeit hoch. So werden hohe Unsicherheiten und Schwankungen verhindert, die einer hochauflösenden Prognose von weit entfernten Störungen innewohnen. Gleichzeitig können die für die Planbarkeit von Instandhaltungsmaßnahmen so wichtigen Vorlaufzeiten von drei bis fünf Tagen oder kalendarisch relevante Zeiträume berücksichtigt werden, z.B. Tage, Wochen oder Monate. Mit der letztlich getroffenen Struktur von zehn Klassen und den optimalen Parametereinstellungen konnte das Potenzial für eine präzise und zuverlässige RUL-Prognose demonstriert werden. Mehr als 50 Prozent aller Störungen werden korrekt und rechtzeitig vorhergesagt, wodurch auch mindestens die Hälfte der Verspätungsminuten (VMin) reduziert werden. Die Analyse der Break Even Situation ergab zudem, dass sowohl kNN als auch SVM so präzise sind, dass ihr Einsatz wirtschaftlich ist. Es wird durch verhinderte Verspätungskosten mehr eingespart, als durch Fehlalarmkosten induziert wird. Wie in Tabelle 6-6 noch einmal zusammengefasst ist, erfüllt die Vorhersage damit alle in Kapitel 4 formulierten Anforderungen – zumindest theoretisch.

Nummer	Kurzbeschreibung	Absicht	Status
1.	Vorhersage in Form einer RUL	Planbarkeit	✓
2.	Vorhersage mit TP Rate größer 0.5 für die Klasse der RUL von fünf oder weniger Tagen	mehr Pünktlichkeit in DE	✓
3.	Summe der VMin von TP in der Klasse mit der RUL von fünf oder weniger Tagen ist größer 50%	mehr Pünktlichkeit in DE	✓
4.	Erreichen der Break Even Situation	Wirtschaftlichkeit	✓
5.	Einsatz von so wenig Sensoren wie möglich	Wirtschaftlichkeit	✓
6.	Nutzung von Sensoren, die für den regulären Bahnbetrieb zugelassen sind	Wirtschaftlichkeit	✓
7.	Verwendung von Weichen mit verschiedenen Eigenschaften	Allgemeingültigkeit	✓
8.	Verwendung von mindestens zehn Weichen	Allgemeingültigkeit	✓
9.	Verwendung von häufig im Netz der DB AG verbauten Weichen	Anwendbarkeit in DE	✓
10.	Erhebung von realen Mess- und Störungsdaten	Anwendbarkeit in DE	✓
11.	Datenbasis über den Zeitraum von mindestens einem Jahr	Anwendbarkeit in DE	✓

Tabelle 6-6: Status (✓ = erfüllt, ? = offen, x = nicht erfüllt) aller Anforderungen an die zu entwickelnde Prognose mit der Rohdatenbasis, der Klassenstruktur und den Verfahren kNN und SVM

Diese Einschränkung muss vorgenommen werden, da die abschließende Anwendung der Prognose unter realen Bedingungen die anfänglichen Erfolge nicht bestätigen konnte. Im Unterschied zur anfänglich datumsunabhängigen Teilung bei Trainings- und Testmenge wurde für die realen Bedingungen ein genaues Datum zur Partitionierung verwendet. Alle Störungen vor diesem Datum dienten dem Training, alle danach dem Test. Die Prognosegenauigkeit lag dabei deutlich unter den Ergebnissen der datumsunabhängigen Prognose. Eine Überanpassung kann als Ursache dafür ausgeschlossen werden, da sich die datumsabhängigen Ergebnisse signifikant verbesserten, wenn den Trainingsdaten auch nur ein Prozent der zu prognostizierenden Daten beigemischt wurde. Viel wahrscheinlicher ist daher, dass bei der datumsabhängigen Partitionierung einfach

nicht mehr genügend Referenzen vorhanden sind, um die verschiedenen Entwicklungsphänomene je Störung und Bauartcluster präzise zu erlernen. Es lässt sich auch vermuten, dass unbekannte Langzeitphänomene von Alterung und Abnutzung sowie abrupte Veränderungen durch Wartungsmaßnahmen ebenfalls nicht mehr mittrainiert wurden. Indizien dafür lieferte eine individuelle Analyse je Störung. Wenn eine einzelne zukünftige Störung mittels einer einzelnen vergangenen Störung vorhergesagt wurde, gab es einige Paarungen mit guten Prognoseerfolgen. Meist war dies der Fall, wenn innerhalb des Bauartclusters dieselbe Störungsursache vorlag und zwischen beiden Störungen nicht mehr als vier Monate vergingen. Allerdings bleiben dies Einzelfälle. Denn ohne deutlich umfangreichere Daten, die für jeden Bauartcluster oder besser für jede Weiche mindestens zwei Störungen derselben Ursache beinhalten, lässt sich dies nicht verallgemeinern. Solche Daten sind aber derzeit nicht verfügbar. Deshalb kann innerhalb der vorliegenden Arbeit nicht abschließend bewiesen werden, dass die entwickelte RUL-Prognose auch anwendungsreif ist. Sie ist allerdings präzise, verlässlich und wirtschaftlich.

7 Zusammenfassung und Ausblick

Im Zugverkehr verursachen Verspätungen jährlich einen enormen volkswirtschaftlichen Schaden in nicht bezifferbarer Höhe. Gleichzeitig zählt Unpünktlichkeit zu den größten Ärgernissen für Fahrgäste. Eisenbahnunternehmen, allen voran die DB AG, leiden nicht nur unter den negativen Folgen für ihr Image, sondern auch unter den Kosten der Verspätungen und der Instandhaltung zur Behebung von Störungen. Die negativen Auswirkungen ließen sich nur verhindern, wenn die Störung rechtzeitig prognostizierbar wäre, um effektive und effiziente Gegenmaßnahmen zu planen und auszuführen. In der vorliegenden Arbeit wurde ein Verfahren entwickelt, das Störungen an Weichen präzise und verlässlich vorhersagt und zudem die Instandhaltungskosten senkt. Sowohl darin als auch in puncto Realitätsnähe und Umfang der verwendeten Daten hebt sich diese Arbeit deutlich von anderen Ansätzen zur Weichenzustandsdiagnose oder -prognose ab. Anstelle von Labordaten zu einer oder wenigen Weichen aus einem kurzen Zeitraum wurden hier mehrere hunderttausend Datensätze aus dem zweijährigen Realbetrieb von 29 unterschiedlichen Weichen zugrunde gelegt. Dabei waren heterogene, umfangreiche Datenquellen zu berücksichtigen, versteckte Zusammenhänge zu identifizieren und geeignete Evaluationsmethoden zur Ergebnisbewertung zu erarbeiten. Deshalb orientierte sich die datengetriebene Entwicklung der Prognose an der Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases) (KDD). Im Entwicklungsverlauf sind an vielen Stellen neue Erkenntnisse und Ansätze entstanden, um die zu Beginn der Arbeit gestellten Forschungsfragen zu beantworten. Im Folgenden wird dies noch einmal zusammenfassend betrachtet.

7.1 Zusammenfassung der Ergebnisse

Aus der detaillierten Betrachtung bestehender Ansätze für die Störungsprognose bei Weichen in Abschnitt 2.4 wurde zwar ersichtlich, dass zu dieser Fragestellung bereits einige Arbeiten verfügbar sind. Doch es wurde gezeigt, dass diese Arbeiten zahlreiche, wichtige Punkte offen lassen. Es mangelt insbesondere an einer Vorhersage der verbleibenden Nutzungsdauer (engl. Remaining Useful Life) (RUL), so dass eine Störung rechtzeitig durch eine geplante Instandhaltungsmaßnahme verhindert werden kann. Was „*rechtzeitig*“ bedeutet, wurde im Rahmen dieser Arbeit erstmals mit dem potenziellen Anwender DB Netz AG definiert. Demnach muss eine Störung für die Feinplanung drei bis fünf Tage vor ihrem Eintreten bekannt sein.

Auf Basis dieser Aussage wurde bewiesen, dass das kommerzielle Weichendiagnosesystem allein nicht ausreicht, um eine Vorhersage zu ermöglichen (siehe Abschnitt 5.1). Selbst der im Rahmen dieser Forschung erstmals verfolgte Ansatz, mehrere Alarmer aus der Stellstromüberwachung zu kombinieren, führte nicht zum Erfolg. Insbesondere ist die so formulierte Vorhersage zu ungenau, um wirtschaftlich zu sein. Dies wurde anhand der eigens entwickelten Methode zur Break Even Situation belegt. Sie erweitert die gängige Performancemetrik des Receiver Operating Characteristic (ROC)-Graphs um eine weitere Dimension für die Störungshäufigkeit. In Verbindung mit den Fehlalarmkosten wegen unnötiger Instandhaltungsmaßnahmen und dem Nutzen einer verhinderten Störung, entsteht eine Ebene, die eine wirtschaftliche Performance von einer unwirtschaftlichen trennt. So wird ersichtlich, unter welchen Bedingungen die Prognose mehr Kosten einspart als sie zusätzlich auslöst – nicht nur bei Weichenstörungen, sondern allgemein.

Nichtsdestotrotz stellen die Stellstrommessungen des Diagnosesystems zentrale Größen für die Prognose bereit. Sie allein sind allerdings nicht ausreichend. Deshalb analysierte diese Arbeit in Abschnitt 5.2 strukturiert, welche Daten gegebenenfalls für die Prognose zu ergänzen sind. Erstmals wurden das Verhalten und die Volatilität der Stellstrommesswerte anhand weiterer Parameter identifiziert bzw. quantifiziert. Der Einfluss des Klimas auf die Messungen beruht auf der Temperatur und der Luftfeuchtigkeit, weil sie Reibungswiderstände bei der Zungenbewegung verändern. So kann Schwergang entstehen, der jedoch nicht auf eine Störung hindeutet. Beide Größen müssen daher in die Prognose einbezogen werden. Der Einfluss, welchen die Bauarteigenschaften auf die Messwerte ausüben, wurde mit einer eigenen, neuartigen Methode analysiert. Dies ist notwendig, da die Voraussetzungen für herkömmliche Analyseverfahren nicht gegeben sind, etwa unabhängige, normalverteilte Attribute mit gleichen Varianzen. Bei der entwickelten Heuristik wurden alle möglichen Cluster mit Weichen ähnlicher Bauarteigenschaften gebildet. Die anschließende Klassifikation der Bauartclusterzugehörigkeit anhand der Stellstromwerte erreichte dabei die besten Werte, wenn die Weichen gemäß des Weichentyps, der Verschlussart und der Zungenrollvorrichtung gruppiert wurden. Im Ergebnis werden von diversen Bauarteigenschaften nur diese drei genutzt, um Weichen zu gruppieren. Statt Weichen individuell zu behandeln, wird die Prognose auf diese Gruppen ausgerichtet, weil Weichen innerhalb der Gruppen ähnliche Stellstromverläufe besitzen. Auf diese Weise kann auch der Zustand neuer Weichen prognostiziert werden, für die noch keine historischen Daten vorliegen. Auf der Grundlage der analysierten Zusammenhänge wurde die Rohdatenbasis aus verschiedenen Quellen selektiert. Die ursprünglich rund 100 Attribute sind auf 28 relevante eingegrenzt worden: einige Stellstromattribute, die Temperatur, die Luftfeuchtigkeit, der Weichentyp, die Verschlussart und die Anzahl der Zungenrollpaare sowie einige Attribute der SAP-Dokumentation von Störungen. Letztere bilden die einzige Referenz für den tatsächlichen Zustand.

Kapitel 6 erläutert die eigentliche Prognose der Störungen. Dieser liegt die Idee zugrunde, die stetige RUL in diskrete Klassen zu transformieren. Das ist zwar ein ungewöhnlicher Ansatz, bringt aber einige Vorteile mit sich. Zum einen wird die Vorhersage zu einem Klassifikationsproblem multipler Klassen, wofür im Data Mining zahlreiche leistungsfähige Verfahren existieren. Zum anderen kann die Klassenstruktur, also die Anzahl der Klassen und das jeweils umschlossene Intervall an RUL-Werten, so gewählt werden, dass die Prognose präzise und verlässlich ist sowie den Anforderung an die Planbarkeit genügt. Weil diese Diskretisierung neu ist und es mit RUL-Prognose bei Weichen noch keine Erfahrungen gibt, liegen nur wenige Anhaltspunkte vor, wie eine Klassenstruktur gestaltet sein sollte. Die vorliegende Arbeit untersuchte deswegen auch den Prognoseerfolg bei verschiedenen Klasseneinteilungen. Daraus ergaben sich einige wichtige Grundregeln für die Diskretisierung. So sind ungefähr zehn Klassen eine gute Wahl, wenn die Granularität passend gewählt wird. D.h. große RUL-Werte oberhalb von 14 bis 28 Tagen sind in Klassen mit weitem Intervall zusammenzufassen. So werden die hohe Unsicherheit und Schwankungen verhindert, die einer hochauflösenden Prognose von weit entfernten Störungen innewohnen. Mittlere Klassengrößen sind im Bereich von 3 bis 28 Tagen zu wählen. Für die Zeitpunkte kurz vor der Störung sollten nur sehr wenige Klassen gebildet werden, die im Intervall von 12 bis 24 Stunden liegen. Dadurch ist zwar die RUL-Auflösung geringer, aber die Prognose zuverlässiger. Hierbei werden die für die Planbarkeit von Instandhaltungsmaßnahmen so wichtigen Vorlaufzeiten von drei bis fünf Tagen und kalendarisch relevante Zeiträume berücksichtigt.

Die Arbeit verwendete zehn Klassen (mit den RUL-Stufen mindestens noch 0, 0.125, 1, 3, 5, 14,

28, 61, 122, 244 Tage bis zur Störung). Die Zugehörigkeit zu diesen Klassen wurde anhand der Messattribute gelernt, um anschließend eine Testdatenmenge zu klassifizieren bzw. deren RUL zu prognostizieren. Weil die Performance von dafür infrage kommenden Verfahren stark von ihrer Parametrierung abhängt, wurden jene Einstellungen herausgearbeitet, bei denen der Erfolg maximal ist. Diese systematische Suche nach passenden Verfahren und ihren bestmöglichen Einstellungen stellt einen wichtigen Forschungsbeitrag für die Weichenzustandsprognose dar, analog zur Lösung unzähliger Data Mining Problemen in der Biologie, der Medizin, der Wirtschaft, etc. Mit k-nächsten-Nachbarn (engl. k-Nearest-Neighbor) (kNN) und der Stützvektormaschine (engl. Support Vector Machine) (SVM) haben sich zwei Klassifikationsverfahren herauskristallisiert, die die Ziele dieser Arbeit erfüllen und somit nachweislich die Forschungsfragen beantworten.

1. Mit beiden Verfahren lassen sich mindestens **50 Prozent der Weichenstörungen rechtzeitig vorhersagen**. In den Klassen mit einer RUL von fünf oder weniger Tagen erreicht kNN eine Rate an richtig klassifizierten Messungen (TP Rate) von 0.987, und SVM kommt sogar auf 0.995. Selbst mit den 35 Störungen, welche aus der Rohdatenbasis ausgeschlossen wurden, weil sie nicht über diese Daten detektierbar sind, lassen sich 63 Prozent verhindern ($63 \cdot 0.987/98 \approx 63$).
2. Weil nur ein Durchschnittswert an Verspätungsminuten (VMin) je Störung verfügbar war, kann allein anhand der 63 Prozent geschlussfolgert werden, dass die Verfahren den Netzbetreiber in die Lage versetzten, auch **50 Prozent der durch Weichen verursachten Verspätungsminuten zu reduzieren**. Wird dies mit den minimalen Kosten von 15 Euro je VMin multipliziert, ergibt sich in Bezug auf das Jahr 2015 ein Einsparpotenzial von 20 Millionen Euro bei den Betriebserschwerungskosten und 2.5 Jahre mehr Pünktlichkeit.
3. Beide Verfahren arbeiten wirtschaftlich, denn sie tragen auch zur **Senkung der Instandhaltungskosten** bei. Sie erreichen in nahezu allen Situationen eine wirtschaftliche Performance, wie anhand der Break Even Situation belegt wurde.
4. Da die Daten, auf denen die Prognose beruht, von einem marktgängigen Diagnosesystem und von frei zugänglichen Quellen stammen, ist eine **minimale Sensorausrüstung** gegeben, die **vollständig rückwirkungsfrei auf die Leit- und Sicherungstechnik** ist. Dadurch wird sichergestellt, dass der Netzbetreiber bestehende Anlagen einfach nachrüsten kann, ohne dass eine Zulassung durch das Eisenbahn-Bundesamt notwendig wird.

Obwohl die Ziele erreicht werden, kann das Potenzial der entwickelten Störungsvorhersage nicht umgehend realisiert werden. Das ergab die abschließende Betrachtung der Prognose unter realen Bedingungen. Im Unterschied zum anfänglichen Vorgehen wurde dafür die Trainings- und Testmenge datumsabhängig geteilt, mit deutlich geringerer Vorhersagegenauigkeit. Vermutlich sind für die verschiedenen Entwicklungsphänomene je Störung und Bauartcluster nicht mehr genügend Referenzen vorhanden. Unbekannte Langzeitphänomene von Alterung und Abnutzung sowie abrupte Veränderungen durch Wartungsmaßnahmen wurden nicht mehr mitgelernt, können aber die Messwerte mit der Zeit verändert haben. Indizien dafür lieferte eine individuelle Analyse je Störung. Wenn eine einzelne zukünftige Störung mittels einer einzelnen vergangenen Störung vorhergesagt wurde, gab es einige Paarungen mit guten Prognoseerfolgen. Doch dies bleiben Einzelfälle. Für eine Verallgemeinerung sind erheblich umfangreichere Daten notwendig. Diese sollten mindestens zwei Weichen mit genau denselben Bauarteigenschaften enthalten, die jeweils mindestens zweimal dieselbe Störungsursache aufweisen. Bestenfalls sind beide Weichen

gleich alt und betrieblich gleich belastet. Dann könnte der Einfluss von nicht erfassten Größen vielleicht ausgeschlossen werden. Solche Daten sind aber derzeit nicht verfügbar.

Die Ergebnisse dieser Arbeit sind dennoch ein wichtiger Beitrag, um den Schienenverkehr leistungsfähiger und kostengünstiger zu gestalten. Die präzise, verlässliche und wirtschaftliche Vorhersage von Weichenstörungen kann der DB AG hohe Betriebserschwerniskosten ersparen. Aber vor allem kann sie den jährlichen volkswirtschaftlichen Schaden reduzieren und die Pünktlichkeit erhöhen und somit die Attraktivität für den Reisenden steigern.

7.2 Ausblick auf weiterführende Forschung

Während der Entwicklung der Prognose konnten wichtige Erkenntnisse gewonnen und Fragen beantwortet werden. Es wurden auch weiterführende oder angrenzende Fragen neu aufgeworfen. Sie sind im Folgenden kurz beschrieben.

Es wurde schon angesprochen, dass eine umfangreichere Datenbasis notwendig ist, um die Prognoseergebnisse unter realen Bedingungen zu validieren. Daneben ist auch eine Erweiterung um Daten der betrieblichen Belastung und von Instandhaltungsmaßnahmen lohnenswert. Analog zu den klima- und bauartbedingten Einflussfaktoren, könnten so auch die belastungs- und instandhaltungsbedingten Einflüsse quantifiziert werden. Dadurch würde eine Prognose allgemeingültiger und wahrscheinlich auch unter Realbedingungen verlässlich. Außerdem könnten dann auch Störungsursachen prognostiziert werden, die über den Stellstrom nicht zu detektieren sind, beispielsweise am Radlenker oder an den Weichenzungen.

Für das neu entwickelte Analyseverfahren, welches den Zusammenhang zwischen Bauarteigenschaften und Stellstrommessungen identifiziert, ist eine von der Thematik der vorliegenden Arbeit unabhängige Diskussion wünschenswert. Ziel der zukünftigen Forschung sollte es sein, zu ergründen, ob es sich als statistisches Standardverfahren eignet. Die Herausforderung besteht dabei in einer allgemeinen, wissenschaftlichen Betrachtung der Eigenschaften und der Leistungsfähigkeit des Verfahrens. Der Schwerpunkt läge daher bei grundlegenden Fragen um das Clustering und die Klassifikation als Basis des Analyseverfahrens.

Als Referenz für den tatsächlichen Zustand wurde die SAP-Dokumentation der Störungen genutzt. Im Zuge dessen wurden Störungen durch manuelle Auswertung der Meldungstexte ausgeschlossen, deren Ursachen sich nicht anhand des Stellstroms offenbaren. Forschung zur automatischen Ursachenanalyse mittels Text Mining ist deshalb interessant, weil so die Referenzen für die Lernbasis automatisiert und kontinuierlich erweitert werden könnten. Davon abgesehen wird eine manuelle Analyse bei deutlich mehr Störungen zu aufwendig. Die Störungsdiagnose kann hingegen dazu eingesetzt werden, die Ursachen automatisch zu dokumentieren. So wird die Dokumentation selbst verlässlicher und einheitlicher gegenüber den heutigen manuellen Freitexteinträgen.

Weil die Prognose auf Wirkleistungswerten aus SIDIS W basiert, stellt sich die Frage, wie sich alternative Größen auswirken, zum Beispiel die Stromstärkemessungen des VAE Roadmasters. Zumal dieses System seit 2014 auch bei der DB Netz AG vermehrt angeschafft wird. Außerdem verspricht die Nutzung der gesamten Messung während eines Weichenumlaufs einen deutlich

höheren Informationsgehalt gegenüber den bisher nutzbaren Durchschnittswerten. Komplette Umlaufmessungen könnten direkt oder mit daraus abgeleiteten, charakteristischen Merkmalen verwendet werden, wie weitergehende Forschungen zur Weichenzustandsdiagnose in Schenkendorf und Böhm [SB14] oder Schenkendorf et al. [SLB14] zeigen.

Einer der spannendsten Forschungsaspekte ergibt sich, wenn die Prognose tatsächlich einsetzbar ist. Dann kann mit der RUL die Instandhaltungsstrategie und der Einsatz von Ressourcen optimiert werden. Wer hat wann Dienst? Welche Werkzeuge und Maschinen sind wann und wo im Einsatz? Dies sind nur die naheliegenden Fragen. Es ergibt sich dann eine Wechselwirkung mit den RUL-Klassen, für deren Struktur es bisher kaum Anhaltspunkte gibt. Das könnte sich im Zuge einer Optimierung ändern, da aus dem erweiterten Planungshorizont, dem Optimierungsziel und den notwendigen Eingangsgrößen Anforderungen an die RUL-Klassen entstehen. Aber auch generell sollten die Bedürfnisse einer Instandhaltung an die RUL-Klassen genauer untersucht werden. Insgesamt spielt die Frage nach einer netzweiten Gesamtoptimierung, wie etwa in der Dissertation von Gutsche [Gut10] ausgearbeitet, eine wichtige Rolle. Dabei ließen sich mehrere Maßnahmen verschiedener Infrastrukturelemente zusammenlegen. Dies setzt allerdings voraus, dass auch der Zustand der anderen Elemente prognostiziert werden kann.

Einige dieser Punkte werden aktuell und zukünftig erforscht (unter anderem am Institut für Verkehrssystemtechnik), nicht zuletzt weil das Thema einer präzisen, verlässlichen und wirtschaftlichen Zustandsprognose auch für andere Verkehrsträger und -infrastrukturen bedeutsam ist.

A Mathematische Beschreibungen von Performance Metriken

A.1 Mathematische Beschreibung von Änderungen der Konfusionsmatrix bei Invarianzeigenschaften von Metriken

Die Abbildung 3-7 zeigt die Felder einer Konfusionsmatrix für ein binäres Klassifikationsproblem. Performance Metriken, die aus einer Konfusionsmatrix berechnet werden, können unterschiedlich auf bestimmte Änderungen in der Konfusionsmatrix reagieren (vgl. Abschnitt 3.3). Die Änderungen sind im Folgenden mathematisch beschrieben.

Die Konfusionsmatrix \mathbf{C} :

$$\mathbf{C} = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (\text{A.1})$$

Eine Performance Metrik für einen binären Klassifikator ist eine Funktion der Konfusionsmatrix.

$$f: \mathbf{C} \rightarrow \text{Metrik} \quad f(\mathbf{C}) = f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) \quad (\text{A.2})$$

Eine Metrik ist invariant gegenüber dem **Tausch der Klassen** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}\right) \quad (\text{A.3})$$

Eine Metrik ist invariant gegenüber der **Änderung von TP** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} TP' & FN \\ FP & TN \end{bmatrix}\right) \quad \text{mit } TP \neq TP' \quad (\text{A.4})$$

Eine Metrik ist invariant gegenüber der **Änderung von FN** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} TP & FN' \\ FP & TN \end{bmatrix}\right) \quad \text{mit } FN \neq FN' \quad (\text{A.5})$$

Eine Metrik ist invariant gegenüber der **Änderung von FP** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} TP & FN \\ FP' & TN \end{bmatrix}\right) \quad \text{mit } FP \neq FP' \quad (\text{A.6})$$

Eine Metrik ist invariant gegenüber der **Änderung von TN** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} TP & FN \\ FP & TN' \end{bmatrix}\right) \quad \text{mit } TN \neq TN' \quad (\text{A.7})$$

Eine Metrik ist invariant gegenüber einer **skalaren Änderung der Konfusionsmatrix** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} \alpha TP & \alpha FN \\ \alpha FP & \alpha TN \end{bmatrix}\right) \quad \text{mit } \alpha \neq 1 \quad (\text{A.8})$$

Eine Metrik ist invariant gegenüber einer **Änderung in der Klassenhypothese** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} \alpha TP & \beta FN \\ \alpha FP & \beta TN \end{bmatrix}\right) \quad \text{mit } \alpha \neq \beta \quad (\text{A.9})$$

Eine Metrik ist invariant gegenüber einer **Änderung in den Klassen** wenn gilt:

$$f\left(\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}\right) = f\left(\begin{bmatrix} \alpha TP & \alpha FN \\ \beta FP & \beta TN \end{bmatrix}\right) \quad \text{mit } \alpha \neq \beta \quad (\text{A.10})$$

A.2 Formeln für die Berechnung ausgewählter Metriken

Im Folgenden sind die Formeln zur Berechnung ausgewählter Metriken zur Bewertung einer binären Klassifikation angegeben. Basis dafür ist die Konfusionsmatrix **C** (siehe Abbildung 3-7 und Gleichung (A.1)).

relative Genauigkeit (engl. Accuracy) (Acc):

$$Acc = \frac{(TP + TN)}{(TP + FN + FP + TN)} \quad (\text{A.11})$$

Sensitivität (engl. Sensitivity, Hit Rate oder Recall):

$$Recall = \frac{TP}{TP + FN} = TP_r \quad (\text{A.12})$$

Relevanz (engl. Precision):

$$Precision = \frac{TP}{TP + FP} \quad (\text{A.13})$$

Spezifität (engl. Specifity):

$$Specitivity = \frac{TN}{FP + TN} = 1 - FP_r \quad (\text{A.14})$$

F-Wert (engl. F-score):

$$F - score = \frac{(1 + \beta^2) \cdot (Precision \cdot Recall)}{\beta^2 \cdot Precision + Recall} = \frac{(1 + \beta^2) \cdot TP}{(\beta^2 \cdot (TP + FN) + TP + FP)} \quad (\text{A.15})$$

mit $\beta > 0$

falsch Positive (engl. False Positives) (FP) Rate:

$$FP_r = \frac{FP}{FP + TN} \quad (\text{A.16})$$

richtig Positive (engl. True Positives) (TP) Rate:

$$TP_r = \frac{TP}{TP + FN} \quad (\text{A.17})$$

Der Receiver Operating Characteristic (ROC)-Graph wird aus der FP Rate und der TP Rate gebildet. Die meisten Klassifikatoren ermitteln bei der Klassifikation für jedes Objekt einen Klassenzugehörigkeitswert (vgl. [Faw03]). Außerdem wählt der Klassifikator oder der Anwender ein bestimmtes Niveau für die Zugehörigkeitswerte, ab dem die Objekte der einen oder anderen Klasse zugeordnet werden. Die Tabelle A-1 listet für zwei Klassifikatoren die tatsächlichen Klassen und die ermittelten Klassenzugehörigkeitswerte für die positive Klasse auf. Wird z.B. das Niveau auf 1.0 gesetzt, würden die Klassifikatoren nur Objekte als positiv klassifizieren, wenn der Zugehörigkeitswert größer oder gleich 1.0 ist. In dem Beispiel würden alle Objekte als negativ klassifiziert. Es ergäbe sich eine FP Rate von 0 und eine TP Rate von 0. Wird das Niveau sukzessive gesenkt, ergeben sich die Punkte in Abbildung A-1. Durch die Verbindung der Punkte entsteht die ROC-Kurve. Sie zeigt auch anschaulich, bei welchem Niveau ein Klassifikator besser ist als der andere. In dem Beispiel sind beide Klassifikatoren anfänglich gleich gut. Bis zu einem Niveau von 0.53 ist der Klassifikator 5 besser, da er weiter links oben liegt. Dann aber ist der Klassifikator 6 überlegen. Er erreicht bei einem Niveau von größer oder gleich 0.47 die insgesamt beste Performance, ist also am weitesten links oben.

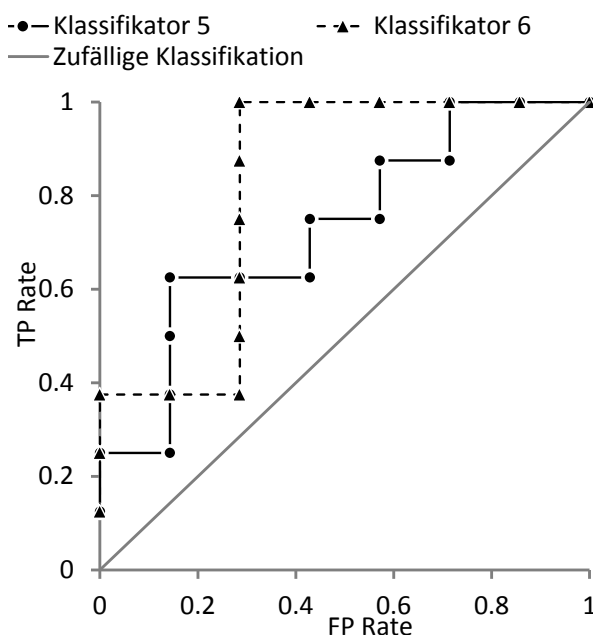


Abbildung A-1: ROC-Kurve von zwei beispielhaften Klassifikatoren

Klassifikator 5		Klassifikator 6	
Tatsächl. Klasse	Wert der Zugehörigkeit zu positiv	Tatsächl. Klasse	Wert der Zugehörigkeit zu positiv
positiv	0.90	positiv	0.80
positiv	0.80	positiv	0.72
negativ	0.70	positiv	0.68
positiv	0.60	negativ	0.53
positiv	0.56	negativ	0.52
positiv	0.54	positiv	0.51
negativ	0.53	positiv	0.50
negativ	0.52	positiv	0.49
positiv	0.50	positiv	0.48
negativ	0.48	positiv	0.47
positiv	0.45	negativ	0.46
negativ	0.40	negativ	0.45
positiv	0.30	negativ	0.40
negativ	0.20	negativ	0.30
negativ	0.10	negativ	0.20

Tabelle A-1: Tatsächliche Klasse der Objekte und ihr Zugehörigkeitswert zu positiv

Fläche unter der ROC-Kurve (engl. Area Under the ROC Curve) (AUC):

$$AUC = \int \hat{F}(\hat{p}|\oplus) d\hat{F}(\hat{p}|\ominus) = \int \hat{F}(\hat{p}|\oplus) \hat{f}(\hat{p}|\ominus) d\hat{p}$$

$\hat{f}(\hat{p}|\oplus) :=$ aus dem Testdatensatz geschätzte Verteilung \hat{p} der Objekte
in der positiven Klasse \oplus (A.18)

$\hat{F}(\hat{p}|\oplus) :=$ kumulierte Verteilung \hat{F} von $\hat{f}(\hat{p}|\oplus)$ verschiedener \hat{p}

B Ergänzende Informationen zur Datenbasis

B.1 Ergänzende Information zur Performance Evaluation von SIDIS W als Störungserdetektion

Abschnitt 5.1.4 nutzt die Metrik des ROC-Graph, um die originäre Funktionsweise von SIDIS W, also die Alarme bei Schwellwertverletzung, zu bewerten. Systematisch wurde untersucht, ob die Kombination mehrerer Alarme eventuell ein besserer Störungsindikator ist. Während die Ergebnisse der jeweiligen FP und der TP Rate in der Tabelle 5-3 aufgelistet sind, werden sie ergänzend im ROC-Graph in Abbildung B-1 dargestellt.

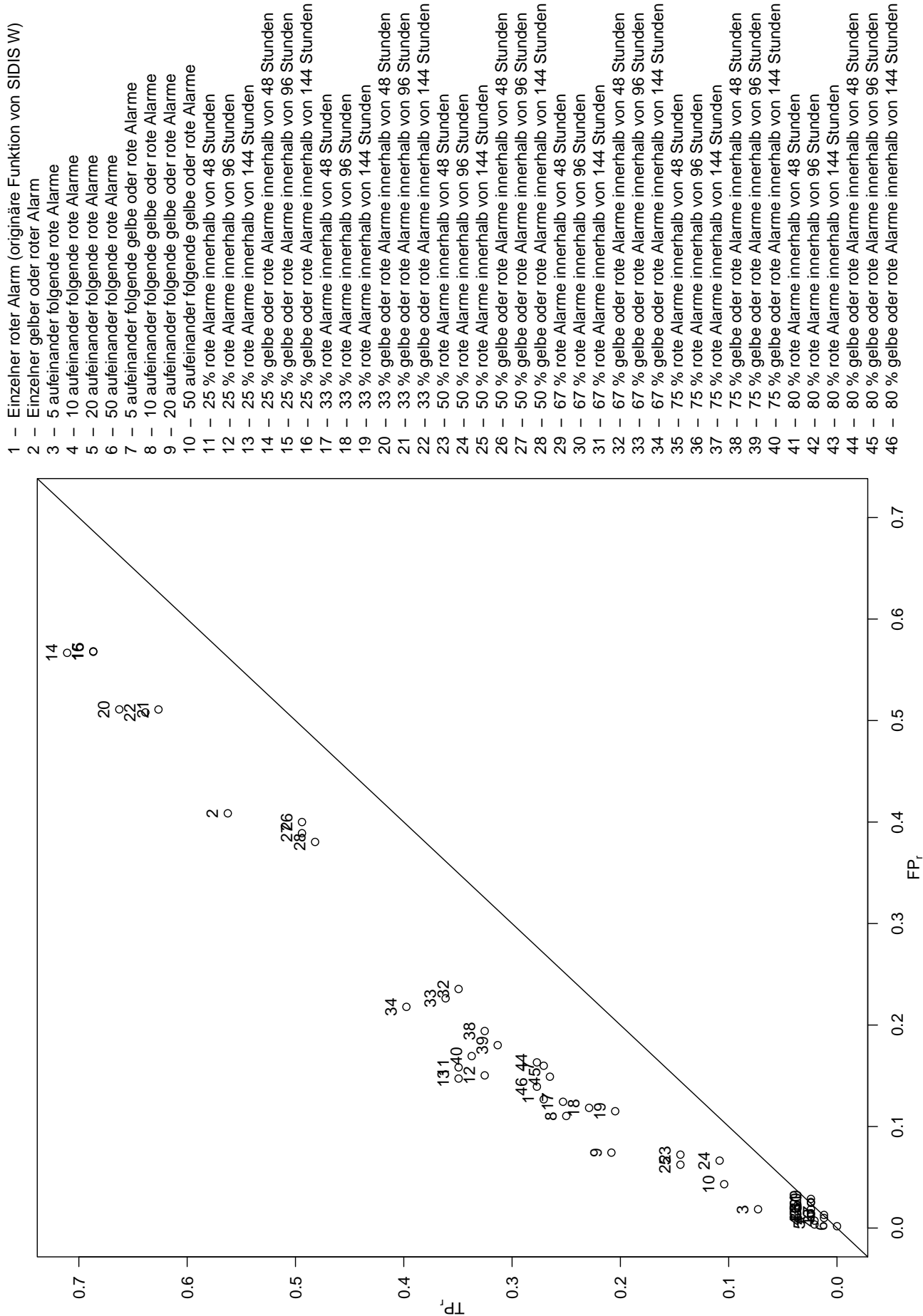


Abbildung B-1: ROC-Graph der verschiedenen Maßstäbe für die Zusammenfassung mehrerer Alarme (Werte gemäß Tabelle 5-3)

B.2 Bedeutung und Verwendung der Attribute in den verfügbaren Daten

Die Grundlage für das entwickelte Prognosemodell bilden verschiedenen Datenquellen, die analysiert wurden, um den Zusammenhang zwischen Störung bzw. RUL abzubilden. Diese Datenquellen sind in Kapitel 5 vorgestellt wurden. Allerdings sind nicht alle gemessenen Eigenschaften (Attribute) der verschiedenen Quellen für das Modell relevant. Um die Ergebnisse auch im Detail nachvollziehen zu können, führen die folgenden Tabellen alle Attribute auf, die in den Rohdaten enthalten sind. Für den Fall, dass Attribute bei der eigentlichen Modellentwicklung deselektiert wurden, sind die entsprechenden Gründe aufgeführt.

Attribut	Bedeutung	Grund für den Ausschluss vom Prognosemodell
AKZ	Anlagenkennziffer (Weichenidentifikation)	
AKZNR	Anlagenkennzifferngruppe für mehrere Weichen	Redundanz zu AKZID
ERFID	Identifikation der Umlaufmessung	Redundanz zur Kombination aus AKZ und ErfDatum
VSaveFlag	unbekannte Bedeutung	sehr kleiner Wertebereich von $\{0, 1, \}$; geringe Varianz; (Keine Vorschläge) über alle Messungen, keine Entropieverringern in Bezug auf Störungen
RingNr	systeminterne Bezeichnung für die angeschlossene Datenleitung	Konstante für alle Messungen
AKZID	Anlagenkennziffer numerisch	Redundanz zu AKZ
InTrend	logischer Wert dafür, ob die Messung bei der systeminternen Trendanalyse berücksichtigt wird oder nicht	sehr kleiner Wertebereich von $\{-1, 0\}$; Wenn Messung nicht wegen Instandhaltung absichtlich negativ; dann Konstante für alle Messungen
ErfDatum	Datum und Uhrzeit der Erfassung	
ErfOK	unbekannte Bedeutung	sehr kleiner Wertebereich von $\{0, 1\}$; geringe Varianz; über alle Messungen; keine Entropieverringern in Bezug auf Störungen
Kurzbeschreibung	Eintrag „Messung automatisch vom System ausgeschlossen“, wenn wegen Instandhaltung absichtlich die Endlage blockiert wurde	Wenn Messung nicht ausgeschlossen, dann sind für alle Messungen keine Werte vorhanden
Richtung	Umlaufrichtung des Motors/Weiche 0=nicht erkannt 1=links 2=rechts	
Strombereich	unbekannte Bedeutung	Konstante 5.0 für alle Messungen
Rzul	Beziehung von Widerstand zu Stromstärke	geringe Varianz; Experimente zeigen keinen Zusammenhang zur Zustandsvorhersage bzw. zeigen Ergebnisverbesserung, wenn das Attribut eliminiert wird.
Pleer	Leistungsdurchschnitt im Antriebsleerlauf	
Plauf1	Leistungsdurchschnitt bei Zungenbewegung im Lauf der ablegenden Zunge	

Tabelle B-1: Verfügbare Attribute in SIDIS W, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell

Tabelle B-1: ... Fortsetzung

Attribut	Bedeutung	Grund für den Ausschluss vom Prognosemodell
Plauf2	Leistungsdurchschnitt bei Zungenbewegung im Lauf beider Zungen	
PverschPeak1	unbekannte Bedeutung	Nicht bei allen Weichen gemessen; tritt normalverteilt auf wenn gemessen; ab dem 28.02.2008 07:17:55 bei keiner Weiche mehr gemessen; keine Entropieverringung in Bezug auf Störungen
Pversch	Leistungsdurchschnitt während des Verschlussvorgangs	
Tumlauf	Zeit des Weichenumlaufs	
TzungAufschl	unbekannte Bedeutung	Mittelwert und Quantile bei 0.0 mit einigen starken Ausreißern, die allerdings nicht im Zusammenhang mit Störungen stehen; Korreliert mit -1 zu TzungUeberd; ab dem 28.02.2008 07:17:55 bei keiner Weiche mehr gemessen;
TzungUeberd	unbekannte Bedeutung	Mittelwert und Quantile bei 0.0 mit einigen starken Ausreißern, die allerdings nicht im Zusammenhang mit Störungen stehen; Korreliert mit -1 zu TzungUeberd; ab dem 28.02.2008 07:17:55 bei keiner Weiche mehr gemessen;
lunsym	Stromasymmetrie, der drei Phasen des Motorstroms	
Pzvorspg1	Leistungsbeitrag durch mechanische Zungenvorspannung	
Pzvorspg2	Leistungsbeitrag durch mechanische Zungenvorspannung	
Tumlauf3p	unbekannte Bedeutung	Mittelwert bei 0.03 (25%-Quantil bei 0.0, 75%-Quantil bei 4.79); nur bei Weichen eines Bahnhofs regelmäßig gemessen, korreliert dort jedoch nur mit 0.3389 mit Tumlauf
Prutschmax	Leistungsspitze beim Durchrutschen der Kupplung im Abschaltpunkt	
TpeakMittel	unbekannte Bedeutung	Konstante 0.0 bei allen Messungen
Temperatur	Temperatur beim Umlauf	wird nicht erfasst, sondern als Konstante von 23°C für alle Messungen angegeben
Xkorrr	Verwendung der korrigierten Wirkleistung 0=nein 1=ja (Die korrigierte Wirkleistung ist ein Wert, der sich aus den ohmschen Verlusten durch den Widerstand der Kabel und des Stators ergibt und von der gemessenen Wirkleistung abgezogen wird.)	Konstante 1.0 bei allen Messungen
Umotor	Motorspannung	
PverschPeakMax	Leistungsspitze am Beginn der Verschlussphase	
RelStellKraft	unbekannte Bedeutung	

Tabelle B-1: Verfügbare Attribute in SIDIS W, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell

Tabelle B-1: ... Fortsetzung

Attribut	Bedeutung	Grund für den Ausschluss vom Prognosemodell
B0F	höchste Alarmstufe (rot, gelb, grün) aller Indikatoren	geringe Aussagekraft über den Weichenzustand
B1F	Alarmstufe des Indikators Tumlauf	geringe Aussagekraft über den Weichenzustand
B2F	Alarmstufe des Indikators lunsym	geringe Aussagekraft über den Weichenzustand
B3F	Alarmstufe des Indikators Pleer	geringe Aussagekraft über den Weichenzustand
B4F	Alarmstufe des Indikators Schwergang über alles (Plauf1, Plauf2, Pverschl)	geringe Aussagekraft über den Weichenzustand
B5F	Alarmstufe des Indikators Plauf2	geringe Aussagekraft über den Weichenzustand
B6F	Alarmstufe des Indikators PverschlPeakMax	geringe Aussagekraft über den Weichenzustand
B7F	Alarmstufe des Indikators Pverschl	geringe Aussagekraft über den Weichenzustand
B8F	nicht verwendet	keine Werte
B9F	nicht verwendet	keine Werte
B10F	nicht verwendet	keine Werte
B0T	Über- oder Unterschreitung oder beides aller Indikatoren	geringe Aussagekraft über den Weichenzustand
B1T	Über- oder Unterschreitung oder beides des Indikators Tumlauf	geringe Aussagekraft über den Weichenzustand
B2T	Über- oder Unterschreitung oder beides des Indikators lunsym	geringe Aussagekraft über den Weichenzustand
B3T	Über- oder Unterschreitung oder beides des Indikators Pleer	geringe Aussagekraft über den Weichenzustand
B4T	Über- oder Unterschreitung oder beides des Indikators Schwergang über alles (Plauf1, Plauf2, Pverschl)	geringe Aussagekraft über den Weichenzustand
B5T	Über- oder Unterschreitung oder beides des Indikators Plauf2	geringe Aussagekraft über den Weichenzustand
B6T	Über- oder Unterschreitung oder beides des Indikators PverschlPeak	geringe Aussagekraft über den Weichenzustand
B7T	Über- oder Unterschreitung oder beides des Indikators Pverschl	geringe Aussagekraft über den Weichenzustand
B8T	nicht verwendet	keine Werte
B9T	nicht verwendet	keine Werte
B10T	nicht verwendet	keine Werte
ErrText	nicht verwendet	keine Werte
AllesDa	unbekannte Bedeutung	sehr kleiner Wertebereich {0,1}; 0.001 % mit 1; keine Entropieverringering in Bezug auf Störungen

Tabelle B-1: Verfügbare Attribute in SIDIS W, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell

Attribut	Bedeutung	Grund für den Ausschluss vom Prognosemodell
Meldender	Bezeichnung der Person oder Rolle, die die Störung gemeldet hat	
Meldung	Meldungsnummer in SAP	
StörBegDatum	Datum und Uhrzeit des Störungsbeginns	
StörEndDatum	Datum und Uhrzeit des Störungsendes	
Auftrag	Auftragsnummer zur der Maßnahme in SAP	ohne Bedeutung für die Modellentwicklung, da nur für Buchhaltung und Organisation relevant
AKZ	Anlagenkennziffer (Weichenidentifikation)	
TechnPlatz	eindeutige SAP-spezifische Identifikation der Anlage, inklusive Ort, Typ und Bauteilen	Redundanz zu AKZ
Beschreibung	Freitext zur Beschreibung der Störung	
Priorität	Priorität, mit der die Störung behoben werden muss	
VerantwArbPl	verantwortlicher Instandhaltungsplaner	ohne Bedeutung für die Modellentwicklung, da nur für Buchhaltung und Organisation relevant
Standort	Standortbezeichnung des Planers	ohne Bedeutung für die Modellentwicklung, da nur für Buchhaltung und Organisation relevant
Planergruppe	Gruppenbezeichnung des Planers	ohne Bedeutung für die Modellentwicklung, da nur für Buchhaltung und Organisation relevant
Ursachentext	Freitext zur Beschreibung der Störungsursache	
Ursachencode	Code für die Ursache	keine Werte
UrsCodeTxt	definierter Text für den Code der Ursache	keine Werte
Codegruppe1	Code für die Ursachengruppe	keine Werte
Codegruppe2	Code für die Ursachengruppe des Objektes	keine Werte
Objektteilcode	Code für die Bauteilgruppe	keine Werte
ObjTGrpText	definierter Text für den Code der Bauteilgruppe	keine Werte
ObjTCodeTxt	definierter Text für den Code des Bauteils	keine Werte
ProblCodeTxt	definierter Text für die Störungsursache	
ProblGrpText	definierter Text für die gestörte Anlagengruppe	
Dauer	Dauer der Störung als Differenz zwischen Beginn und Ende	

Tabelle B-2: Verfügbare Attribute des SAP-Auszugs dokumentierter Störungen an Weichen, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell

Attribut	Bedeutung	Grund für den Ausschluss vom Prognosemodell
WErfDatum	Datum und Uhrzeit der Wetterdatenerfassung	
Feuch in %	Luftfeuchtigkeit	
Temperatur in °C	Lufttemperatur	
Aender	Wertänderung	keine Entropieverringering in Bezug auf Störungen
Luftd in hPa	Luftdruck	
Windr in °	Windrichtung	
Nieder in mm	Niederschlag	
Winds in kn	Windgeschwindigkeit	
Sicht in m	Sichtweite	Starke einfache Korrelation mit Feuch in % (-0.655); polynomiale Regression mittels Temperatur, Feuchtigkeit, Niederschlag als Ersatz möglich
C1	unbekannte Bedeutung	keine Werte
WW	unbekannte Bedeutung	keine Werte
Ch	unbekannte Bedeutung	keine Werte
Cm	unbekannte Bedeutung	keine Werte
Schn in cm	Schneefall	keine Werte bei rund 50 Prozent aller Messungen, Rest konstant bei 0.0

Tabelle B-3: Verfügbare Attribute der Klimadaten, ihre Bedeutung und ggf. Grund ihres Ausschlusses vom Prognosemodell

B.3 Ergebnisse der Korrelationsanalyse zwischen SIDIS W Attributen und Klimadaten

Für die Analyse klimabedingter Einflussfaktoren auf den Weichenzustand oder die Messungen wurde im Abschnitt 5.2.2 die linear Korrelation zwischen einigen Attributen von SIDIS W und Klimadaten untersucht. Da für jede Weiche und Umstellrichtung die Korrelationen gesondert untersucht wurden, entstanden 1102 (29 Weichen mal 2 Richtungen mal 19 Attribute) Datensätze von Korrelationen zu 19 Attributen. Diese können aus Platzgründen weder in dem entsprechenden Abschnitt noch hier im Anhang aufgeführt werden. Die Abbildung 5-7 enthält daher nur den Mittelwert der jeweiligen Korrelationskoeffizienten. Neben diesem Mittelwert ist auch die Schwankung, also die Standardabweichung interessant, da die Korrelation je Weiche und Richtung sehr unterschiedliche sein kann. Die Standardabweichung hilft einzuschätzen, ob die Korrelation grundsätzlich gegeben ist oder nur in einigen Fällen. Die Werte sind in der Tabelle B-4 aufgeführt. Es fällt auf, dass gerade die Temperatur, die Luftfeuchtigkeit und die Sichtweite in einem breiteren Intervall um die Mittelwerte schwanken als die anderen Klimawerte. In Kombination mit den teilweise hohen Korrelationskoeffizienten in Abbildung 5-7 sind diese Klimadaten daher für die weitere Entwicklung des Prognosemodells relevant.

	Rzul	Pleer	Plauf1	Plauf2	Pversch1	Tumlauf	lunsym	Pzvorspg1	Pzvorspg2	Prutschmax	Umotor	Pversch1PeakMax	Feuch in %	Temperatur in °C	Luftd in hPa	Windr in °	Nieder in mm	Winds in kn	Sicht in m
Rzul	0	0.11	0.16	0.46	0.09	0.09	0.10	0.57	0.22	0.12	0.21	0.17	0.02	0.01	0.05	0.02	0.00	0.01	0.02
Pleer	0.11	0	0.33	0.28	0.29	0.32	0.22	0.23	0.27	0.35	0.25	0.24	0.18	0.36	0.13	0.07	0.05	0.10	0.17
Plauf1	0.16	0.33	0	0.21	0.22	0.23	0.19	0.25	0.33	0.33	0.20	0.25	0.22	0.36	0.12	0.10	0.04	0.13	0.23
Plauf2	0.46	0.28	0.21	0	0.20	0.25	0.19	0.21	0.32	0.28	0.22	0.22	0.23	0.31	0.14	0.11	0.06	0.12	0.20
Pversch1	0.09	0.29	0.22	0.20	0	0.21	0.22	0.21	0.32	0.30	0.16	0.24	0.24	0.35	0.11	0.09	0.05	0.11	0.23
Tumlauf	0.09	0.32	0.23	0.25	0.21	0	0.25	0.25	0.28	0.40	0.19	0.29	0.25	0.39	0.10	0.09	0.03	0.13	0.23
lunsym	0.10	0.22	0.19	0.19	0.22	0.25	0	0.17	0.14	0.27	0.26	0.17	0.14	0.20	0.06	0.04	0.03	0.09	0.11
Pzvorspg1	0.57	0.23	0.25	0.21	0.21	0.25	0.17	0	0.27	0.26	0.16	0.24	0.17	0.28	0.10	0.09	0.05	0.10	0.16
Pzvorspg2	0.22	0.27	0.33	0.32	0.32	0.28	0.14	0.27	0	0.31	0.12	0.29	0.23	0.31	0.15	0.10	0.05	0.12	0.22
Prutschmax	0.12	0.35	0.33	0.28	0.30	0.40	0.27	0.26	0.31	0	0.14	0.26	0.27	0.29	0.11	0.08	0.05	0.16	0.21
Umotor	0.21	0.25	0.20	0.22	0.16	0.19	0.26	0.16	0.12	0.14	0	0.18	0.13	0.15	0.04	0.06	0.03	0.15	0.11
Pversch1PeakMax	0.17	0.24	0.25	0.22	0.24	0.29	0.17	0.24	0.29	0.26	0.18	0	0.21	0.30	0.13	0.10	0.05	0.14	0.21
Feuch in %	0.02	0.18	0.22	0.23	0.24	0.25	0.14	0.17	0.23	0.27	0.13	0.21	0	0.21	0.07	0.09	0.05	0.06	0.08
Temperatur in °C	0.01	0.36	0.36	0.31	0.35	0.39	0	0.21	0.30	0.29	0	0.30	0.21	0	0.09	0.08	0.03	0.07	0.17
Luftd in hPa	0.05	0.13	0.12	0.14	0.11	0.09	0	0.09	0.07	0	0.04	0.13	0.07	0.09	0	0.07	0.04	0.05	0.09
Windr in °	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nieder in mm	0.00	0.05	0.04	0.06	0.03	0.05	0.03	0.05	0.05	0.05	0.03	0.05	0.05	0.03	0.04	0.03	0	0.06	0.05
Winds in kn	0.01	0.10	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0	0.09
Sicht in m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabelle B-4: Standardabweichung für das arithmetische Mittel aller linearen Korrelationskoeffizienten je Weiche und Umstellrichtung aus Abbildung 5-7

Im Abschnitt 5.2.2 wurde auf ein Experiment hingewiesen, bei dem die Korrelation einiger SIDIS W Attribute mit der Temperatur dazu benutzt wurde, die Schwankungen zu glätten (vgl. [Böh12a, Böh12b]). Das reduziert zwar die ausgelösten Alarmer, verbessert aber die Genauigkeit von SIDIS W nicht wesentlich, wie der folgende ROC-Graph (Abbildung B-2) zeigt.

- 1 – Einzelner roter Alarm (originäre Funktion von SIDS W)
- 2 – Einzelner gelber oder roter Alarm
- 3 – 5 aufeinander folgende rote Alarme
- 4 – 10 aufeinander folgende rote Alarme
- 5 – 20 aufeinander folgende rote Alarme
- 6 – 50 aufeinander folgende rote Alarme
- 7 – 5 aufeinander folgende gelbe oder rote Alarme
- 8 – 10 aufeinander folgende gelbe oder rote Alarme
- 9 – 20 aufeinander folgende gelbe oder rote Alarme
- 10 – 50 aufeinander folgende gelbe oder rote Alarme
- 11 – 25 % rote Alarme innerhalb von 48 Stunden
- 12 – 25 % rote Alarme innerhalb von 96 Stunden
- 13 – 25 % rote Alarme innerhalb von 144 Stunden
- 14 – 25 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 15 – 25 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 16 – 25 % gelbe oder rote Alarme innerhalb von 144 Stunden
- 17 – 33 % rote Alarme innerhalb von 48 Stunden
- 18 – 33 % rote Alarme innerhalb von 96 Stunden
- 19 – 33 % rote Alarme innerhalb von 144 Stunden
- 20 – 33 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 21 – 33 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 22 – 33 % gelbe oder rote Alarme innerhalb von 144 Stunden
- 23 – 50 % rote Alarme innerhalb von 48 Stunden
- 24 – 50 % rote Alarme innerhalb von 96 Stunden
- 25 – 50 % rote Alarme innerhalb von 144 Stunden
- 26 – 50 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 27 – 50 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 28 – 50 % gelbe oder rote Alarme innerhalb von 144 Stunden
- 29 – 67 % rote Alarme innerhalb von 48 Stunden
- 30 – 67 % rote Alarme innerhalb von 96 Stunden
- 31 – 67 % rote Alarme innerhalb von 144 Stunden
- 32 – 67 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 33 – 67 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 34 – 67 % gelbe oder rote Alarme innerhalb von 144 Stunden
- 35 – 75 % rote Alarme innerhalb von 48 Stunden
- 36 – 75 % rote Alarme innerhalb von 96 Stunden
- 37 – 75 % rote Alarme innerhalb von 144 Stunden
- 38 – 75 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 39 – 75 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 40 – 75 % gelbe oder rote Alarme innerhalb von 144 Stunden
- 41 – 80 % rote Alarme innerhalb von 48 Stunden
- 42 – 80 % rote Alarme innerhalb von 96 Stunden
- 43 – 80 % rote Alarme innerhalb von 144 Stunden
- 44 – 80 % gelbe oder rote Alarme innerhalb von 48 Stunden
- 45 – 80 % gelbe oder rote Alarme innerhalb von 96 Stunden
- 46 – 80 % gelbe oder rote Alarme innerhalb von 144 Stunden

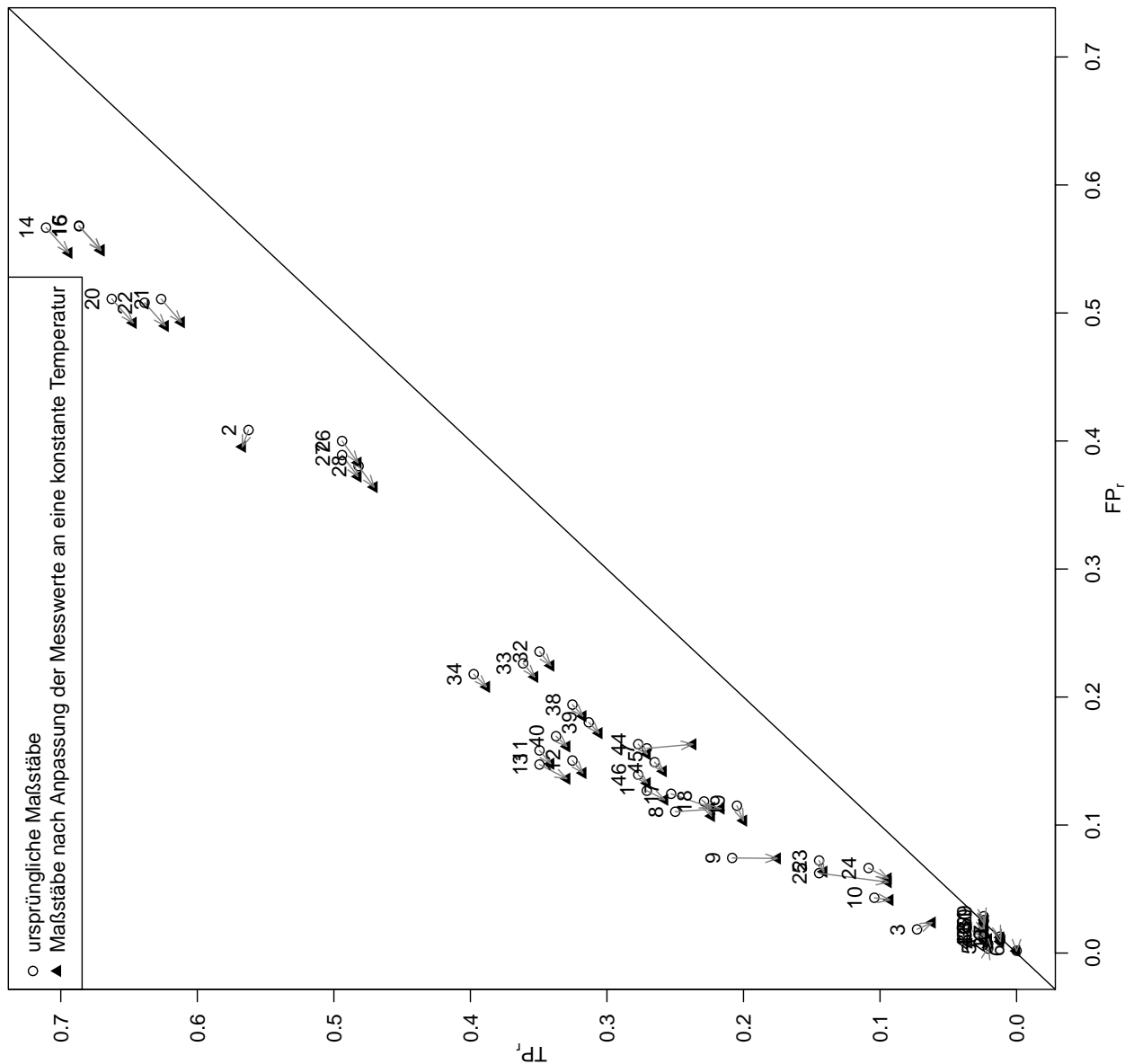
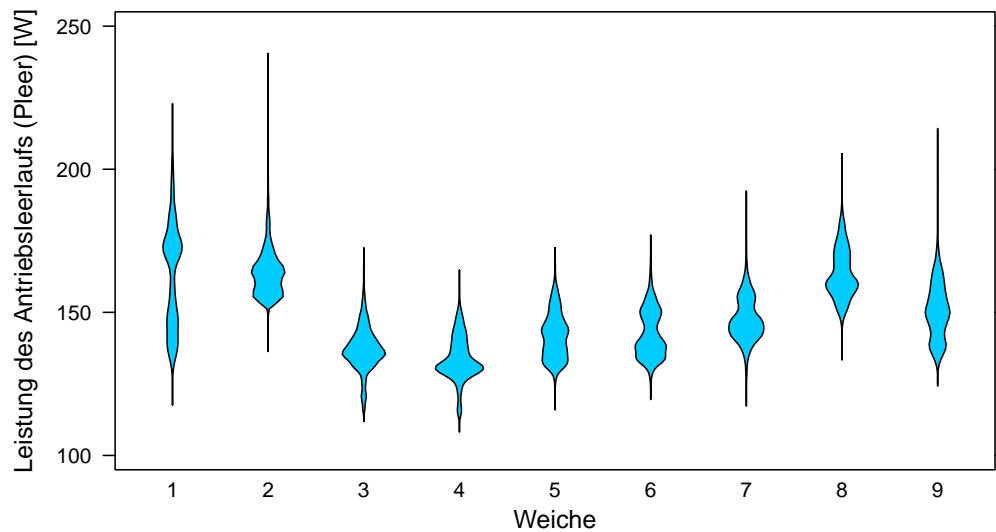


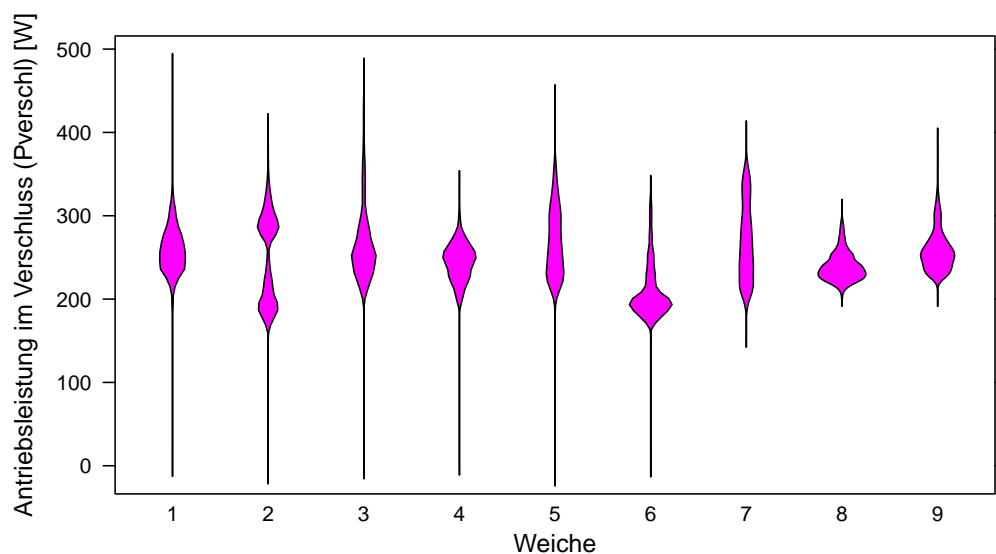
Abbildung B-2: ROC-Graph der verschiedenen Maßstäbe vor und nach der korrelationsgemäßen Temperaturanpassung der Alarme

B.4 Ergebnisse der Analyse zwischen SIDIS W Attributen und Bauarteigenschaften

Im Abschnitt 5.2.3 wurde der Zusammenhang zwischen SIDIS W Attributen und Bauarteigenschaften von Weichen untersucht, um das unterschiedliche Verhalten von Weichen zu erklären. Für die Analyse wurde ein eigenes Verfahren gewählt, weil gängige statistische Analyseverfahren, wie etwa Hotellings T^2 -Test, Wilcoxon-Mann-Whitney-Test, Varianzanalyse (ANOVA) oder Diskriminanzanalyse nicht verwendet werden können. Die mit diesen Verfahren verbundenen Voraussetzungen sind nicht erfüllt, da die Attributwerte weder normalverteilt sind noch gleiche Varianzen aufweisen. Die folgende Abbildung B-3 belegt dies am Beispiel von neun Weichen aus demselben Zeitraum, der auch für die Analyse im Abschnitt 5.2.3 verwendet wurde. Der sogenannte Violinplot zeigt die Verteilung von Werten entlang des jeweiligen Attributes, je breiter die Fläche, desto mehr Messungen besitzen den entsprechenden Wert.



(a) Verteilung der Messwerte im Antriebsleerlauf (Pleer)



(b) Verteilung der Messwerte im AntriebsleerlaufVerschluss (Pverschl)

Abbildung B-3: Violinplot der SIDIS W Attribute Pleer (a) und Pverschl (b) zeigt die unterschiedlichen Verteilungen und Varianzen zwischen Weichen und Attributen

B.5 Von der Zustandsreferenz ausgeschlossene Störungen

Im Abschnitt 5.3 wurde die Rohdatenbasis für die Prognosemodellentwicklung beschrieben. Da in dieser Basis hauptsächlich Daten enthalten sind, die Rückschlüsse über den Weichenzustand anhand des Stellstroms erlauben, wurden bestimmte Weichenstörungen durch Oberbaukomponenten von der Referenz ausgeschlossen. Würden diese Störungen als Referenz bei den Lernverfahren verwendet, würden sie das Ergebnis negativ beeinflussen. Damit der Leser sich selbst ein Bild zu den ausgeschlossenen Störungen machen kann, sind die Störungsdaten aus SAP und die Begründung für den Ausschluss in der Tabelle B-5 aufgelistet. Der Ausschluss erfolgte immer in kombinierter Betrachtung der Standardtexte und der Freitexte. So wurden z.B. nicht grundsätzlich alle Einträge *“unzeitige Auffahrmeldung”* ausgeschlossen, sondern nur dann, wenn aus dem Gesamteintrag ersichtlich, dass keine Maßnahme zur Entstörung stattgefunden hat, die die signifikant Messwerte verändert.

Des Weiteren zeigt die Abbildung B-4 wie sich die Performance von SIDIS W ändert, wenn diese Störungen nicht verwendet werden. Die TP Rate ist zwar besser, aber die Genauigkeit ist dennoch nicht ausreichend.

Wei- che	Stö- rungsID	Beschreibung	Ursachentext	ProblCodeTxt	ProblGrpText	Grund für den Ausschluss von der Re- ferenz
1	98	BCHB WE 301 Zungen- rollen prüfen	Zungenrollen eingestellt	nicht codiert Ein- trag im Textfeld	Schadbild LST Weichen	Zungenroller eingestellt; d.h. keine Wei- chenstörung
1	145	BWKR WE 301 keine Endlage rechts	EHW n. Funktionstüch- tig	keine Endlage	Schadbild LST Weichen	elektrische Weichenheizung defekt, dar- um festgefroren; daher nicht gezählt
2	126	BCHB WE 302 unzeiti- ge Auffahrmeldung »	Verschleiß-Kleineisen verspannt	lose / selbststän- dig gelöst	Schadbild allge- mein	unzeitige Auffahrmeldung durch Kleinei- sen (ELP); daher nicht gezählt
2	171	BCHB Weiche 302 un- zeit.Auffahrmeldung	überprüft, nicht mehr aufgetreten	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
4	50	BCHB, W324 unzeitige Auffahrmeldung, 1x	Rechte Zunge leicht verbogen, muß →	unzeitige Auffahr- meldung	Schadbild LST Weichen	verbogene Zunge nicht über Stellstrom detektierbar; daher nicht gezählt
4	49	BCHB:WE324 Rechte Zunge leicht verbogen,	keine Rückmeldung	Enden verbogen	Schadbild Fahr- bahn/Oberbau	Folge der Störung ID=50; daher nicht gezählt
8	119	BCHB WE 327 Spur/Leitweitenfehler	FSP-Radlenkerbleche eingebaut	Spurenweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
8	132	BWKR-F: W327 Kabel EWH zerissen	Kabel EWH zerissen	gerissen	Schadbild allge- mein	Kabel der elektrische Weichenheizung zerissen; d.h. keine Weichenstörung
8	136	BWKR: W327 r.k.Endlage	Bei Eintreffen Stö.frei.vermutl.mech.	keine Endlage	Schadbild LST Weichen	bei Eintreffen störfrei, vermutliche plötz- liche mechanische Hemmung; daher nicht gezählt
10	4	BCHB WE 329 keine Endlage links	nach Betra keine Heiz- stäbe » werden z.Z	keine Endlage	Schadbild LST Weichen	Heizstab der elektrische Weichenheizung defekt bzw. fehlt; d.h. keine Weichenstö- rung
10	45	BCHB WE329 2 Herz- stückbolzen lose	Herzstückbolzen erneu- ert+festgezogen	abgenutzt	Schadbild allge- mein	Kleineisen (hier Herzstückbolzen) ist lo- se; daher nicht gezählt
10	120	BCHB WE 329 Spur/Leitweitenfehler	FSP-Radlenkerbleche eingebaut	Spurenweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
11	121	BCHB WE 330 Spur/Leitweitenfehler	FSP-Radlenkerbleche eingebaut	Spurenweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
12	88	Bzos BcHB WE 311 keine lila	mech. Hemmung	nicht stellbar	Schadbild LST Weichen	mechanische Hemmung, plötzliches Er- eignis; daher nicht gezählt

Tabelle B-5: SAP-Störungseinträge (Spalten 3-6) und die Begründung für ihren Ausschluss von der Zustandsreferenz bei der Prognosemodellentwicklung

Tabelle B-5: ... Fortsetzung

Wei- che	Stö- rungsID	Beschreibung	Ursachentext	ProblCodeTxt	ProblGrpText	Grund für den Ausschluss von der Re- ferenz
13	124	BCHB WE 332 Leitwei- tenfehler	Radlenkerbleche einge- baut	Spurerweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
15	122	BCHB WE 334 Spur-/Leitweitenfehler	FSP-Radlenkerbleche eingebaut	Spurerweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
15	125	BCHB WE 334 Leitwei- tenfehler	Radlenkerbleche einge- baut	Spurerweiterung	Schadbild Fahr- bahn/Oberbau	Radlenkerbleche eingebaut; d.h. keine Weichenstörung
15	135	BCHB; We 334 Whzg defekt	Heizstäbe gew	reagiert nicht	Schadbild allge- mein	elektrische Weichenheizung defekt; d.h. keine Weichenstörung
16	192	BCHB WE 335 Leitwei- tenfehler	Radlenkerbleche einge- baut			Radlenkerbleche eingebaut; d.h. keine Weichenstörung
17	155	BWKR We337 keine Endlage rechts	40648, WE v.Eis+Schnee befreit	keine Endlage	Schadbild LST Weichen	mechanische Hemmung, plötzliches Er- eignis; daher nicht gezählt
18	17	BZOO, W1211 Ultra- schallfehler	US-Fehler gesichert	nicht codiert Ein- trag im Textfeld	Schadbild Fahr- bahn/Oberbau	Ultraschall-Fehler der Schiene durch Oberbauinspektion ermittelt; daher nicht gezählt
20	131	BZOO WE 1215 unzei- tige Auffm.	kurze am ELP reguliert	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
21	93	BZOO WE 1231 rechts und links k. Endlage	Kontrolle reg.wegen Tempände- rung,4mmProb	keine Endlage	Schadbild LST Weichen	wegen Temperatur reguliert und bei 4mm-Probe absichtlich blockiert; daher nicht gezählt
21	96	BWKR WE 1231 un- z. Auffahrmeldung	nicht zu ermitteln	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
21	97	BWKR WE1231 unzei- tige Auffahrmeldung	nach Betraarbeiten WE überprüft »	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
21	144	BZOO WE 1231 ein- malig k. Endlage	Störung einmalig	keine Endlage	Schadbild LST Weichen	einmalige Störung, vermutliche plötzli- che mechanische Hemmung; daher nicht gezählt
21	156	BZOO, WE 1231 Whz def.	Whz prov wieder herge- stellt,Fdl k.zugrif	Fehlfunktion	Schadbild allge- mein	elektrische Weichenheizung defekt; d.h. keine Weichenstörung

Tabelle B-5: SAP-Störungseinträge (Spalten 3-6) und die Begründung für ihren Ausschluss von der Zustandsreferenz bei der Prognosemodellentwicklung

Tabelle B-5: ... Fortsetzung

Wei- che	Stö- rungsID	Beschreibung	Ursachentext	ProblCodeTxt	ProblGrpText	Grund für den Ausschluss von der Re- ferenz
21	160	BZOO Weiche 1231 1 Heizstab def.,verschm	Heizstab gew.	verwittert / zer- kratzt	Schadbild allge- mein	Heizstab der elektrische Weichenheizung defekt; d.h. keine Weichenstörung
22	83	BWKR WE 1232 r. Zunge muß gerichtet wer.	Keine Rückmeldung	nicht codiert Ein- trag im Textfeld	Schadbild allge- mein	verbogene Zunge nicht über Stellstrom detektierbar; daher nicht gezählt
22	103	BZOO We1232 unzeiti- ge Auffahrmeldung	.überprüft, nicht ermit- telt	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
22	104	BZOO WE 1232 unzei- tiger Auff.	21:15 unterbrochen	unzeitige Auffahr- meldung	Schadbild LST Weichen	unzeitige Auffahrmeldung nicht zu er- mitteln; daher nicht gezählt
22	158	BZOO, WE 1232 Whz def.	Whz prov wieder herge- stellt,Fdl k.zugrif	Fehlfunktion	Schadbild allge- mein	elektrische Weichenheizung defekt; d.h. keine Weichenstörung

Tabelle B-5: SAP-Störungseinträge (Spalten 3-6) und die Begründung für ihren Ausschluss von der Zustandsreferenz bei der Prognosemodellentwicklung

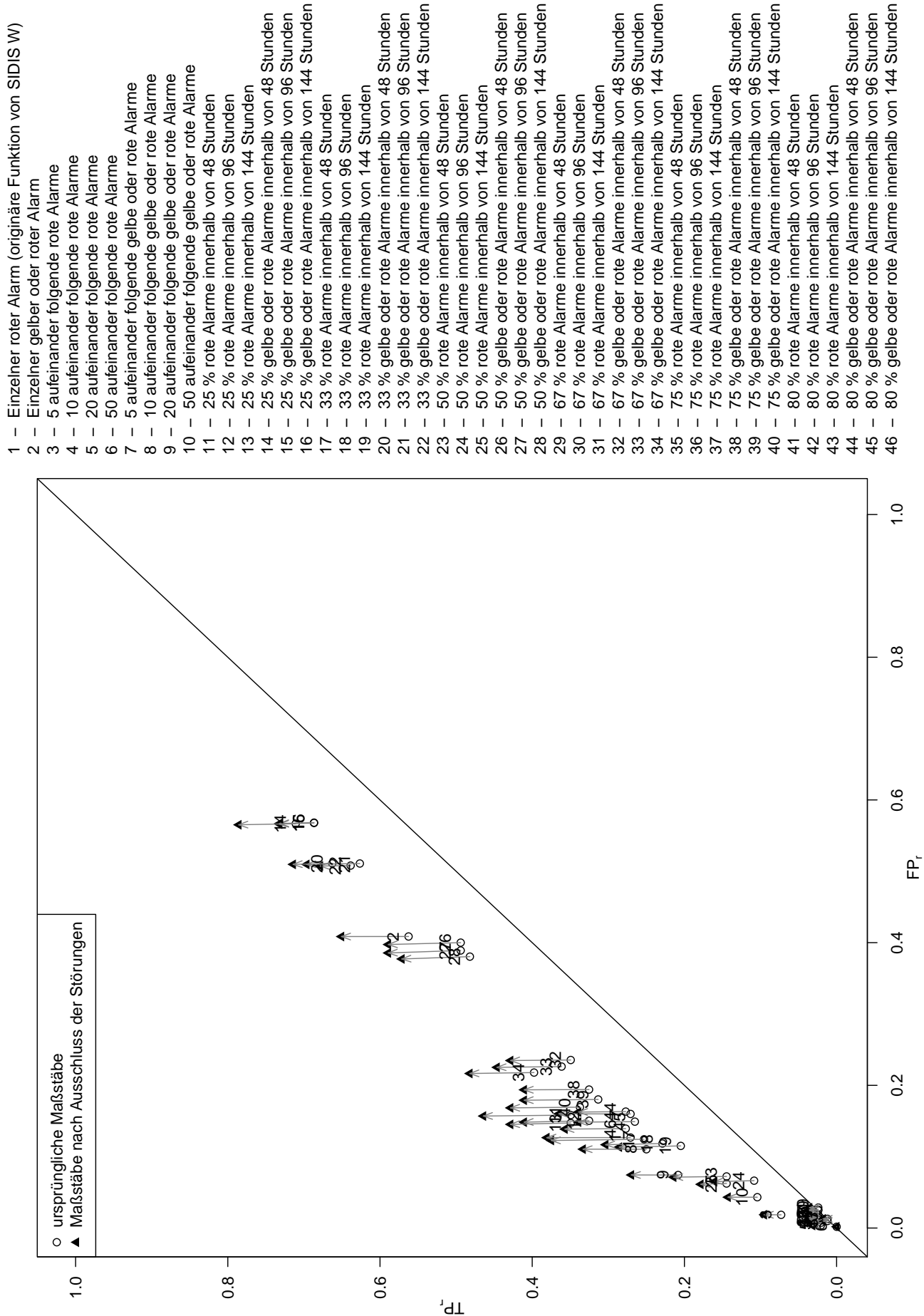


Abbildung B-4: ROC-Graph der verschiedenen Maßstäbe vor und nach dem Ausschluss von Störungen

C Ergänzende Informationen und Auswertungen zur Weichenzustandsprognose

C.1 Prognoseergebnisse ausgewählter Verfahren bei bestmöglicher Parametrierung

Im Abschnitt 6.2.2 wurden die Klassifikationsverfahren kNN, DTree und SVM miteinander verglichen. Dabei wurden die zuvor ermittelten bestmöglichen Parametereinstellungen verwendet. Der Prognoseerfolg wurde mittels Metriken ausgewertet. Die zugrundeliegenden Konfusionsmatrizen werden im folgenden aufgelistet. Die Tabellen C-1, C-2 und C-3 enthalten die Anzahl der Samples je tatsächlicher (C_i) und hypothetischer (prognostizierter) Klasse (C_j). Es ist zu beachten, dass es sich um die Summen der 100 Wiederholungsläufe handelt. In jedem Lauf wurden die Konfusionsmatrizen für die 30-prozentige Testmenge gebildet.

C_j	C_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
C_i	RUL	≤ 0.0625	≤ 0.125	≤ 0.25	≤ 0.5	≤ 1	≤ 2	≤ 4	≤ 8	≤ 16	≤ 32	≤ 64	≤ 128	≤ 256	$\leq \infty$
1		5683	548	289	44	208	25	50	64	28	165	127	82	53	34
2		507	3196	507	49	40	9	0	0	34	6	3	9	4	36
3		266	502	6321	783	57	7	4	0	4	45	1	4	0	6
4		39	35	797	11431	593	9	1	36	0	37	18	4	0	0
5		90	35	91	630	27801	1019	15	8	1	78	68	37	27	0
6		16	5	4	10	1060	55232	1357	53	152	61	30	12	6	2
7		52	0	0	1	1	1352	110608	1285	369	145	109	63	13	2
8		63	0	0	5	60	69	1352	202984	1560	329	106	33	34	5
9		39	6	1	0	1	132	571	1641	397596	2152	362	303	217	79
10		133	7	14	10	47	63	169	384	1852	667426	2278	705	25	87
11		142	5	1	0	18	10	129	147	393	2332	752465	3282	85	191
12		47	10	2	33	49	5	27	10	346	611	3774	989759	3963	664
13		57	10	0	1	6	0	0	35	98	184	205	3694	923292	2718
14		51	0	5	1	0	0	0	1	23	36	49	461	2338	1319035

Tabelle C-1: Konfusionsmatrix der Prognose mittels kNN bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen

C_i	C_j													
	≤ 0.0625	≤ 0.125	≤ 0.25	≤ 0.5	≤ 1	≤ 2	≤ 4	≤ 8	≤ 16	≤ 32	≤ 64	≤ 128	≤ 256	$\leq \infty$
1	417	236	689	306	509	494	387	458	489	1643	870	500	197	205
2	265	705	523	184	270	297	95	300	164	747	394	196	165	95
3	223	77	2406	1317	523	390	254	412	400	685	580	458	83	192
4	102	80	637	5423	1487	666	532	615	553	1091	691	605	223	295
5	34	15	280	1201	16949	3389	1062	830	1303	1768	1338	1017	356	358
6	42	0	43	198	1758	40679	4100	1891	1991	2333	1622	1754	896	693
7	34	1	26	149	396	3310	90657	6717	3585	3604	2525	1585	787	624
8	64	9	120	434	696	1496	5930	173916	9382	6469	3512	2283	1244	1045
9	69	11	56	262	479	1660	3422	9556	349534	19086	9626	4859	2980	1500
10	151	49	239	241	994	1910	2035	5655	17104	595643	30674	10659	4479	3367
11	117	4	144	279	979	1901	1952	3211	7778	28547	664139	36802	7188	6159
12	210	7	175	247	416	1177	1079	2632	4150	11376	33244	902878	33655	8054
13	7	22	85	87	105	627	562	1526	2247	4924	7926	35582	858429	18171
14	13	36	39	36	170	478	664	1223	1182	2875	5589	8010	17221	1284464

Tabelle C-2: Konfusionsmatrix der Prognose mittels DTree bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen

C_j	$C_i \text{ RUL}$	1 ≤ 0.0625	2 ≤ 0.125	3 ≤ 0.25	4 ≤ 0.5	5 ≤ 1	6 ≤ 2	7 ≤ 4	8 ≤ 8	9 ≤ 16	10 ≤ 32	11 ≤ 64	12 ≤ 128	13 ≤ 256	14 $\leq \infty$
1		5363	596	481	9	114	19	0	18	40	255	243	107	10	145
2		599	2735	662	29	92	0	1	0	12	106	11	83	31	39
3		208	670	5840	1120	40	0	0	10	0	50	20	0	0	40
4		1	12	805	11156	936	32	8	37	3	8	2	0	0	0
5		85	22	22	455	28276	714	28	10	29	27	135	53	12	32
6		21	7	6	0	1214	55082	1260	60	200	20	40	0	50	40
7		70	6	1	0	0	1161	111628	846	83	10	93	39	8	55
8		35	0	6	3	68	29	1344	203649	1073	234	109	20	10	20
9		52	0	0	0	0	29	93	1261	399638	1504	266	125	17	115
10		100	0	1	0	2	0	67	85	1345	669468	1609	246	144	133
11		171	0	0	0	32	73	22	81	128	1671	754294	2427	135	166
12		32	9	1	0	58	3	1	34	51	258	1999	994713	1926	215
13		89	7	8	0	7	0	0	32	49	105	236	2357	925776	1634
14		57	0	3	15	2	0	0	0	4	65	25	208	1308	1320313

Tabelle C-3: Konfusionsmatrix der Prognose mittels SVM bei bestmöglicher Parametrierung; Samples je Klasse von 100 Wiederholungsläufen

C.2 Formeln zur Berechnung alternativer RUL-Klassen

Im Abschnitt 6.2.3 wurden alternative Klassen für die RUL gebildet. Für eine bessere Übersichtlichkeit wurde dort darauf verzichtet die formale Definition aufzuführen. Dies sei im folgenden ergänzt. Es wird jeweils die Gleichung (6.1) adaptiert, die für die initiale Klassenstruktur Z_0 genutzt wurde.

- Die drei kleinsten Klassen werden zusammengefasst, so dass unterhalb 0.25 Tagen nicht weiter differenziert wird. Die Struktur wird als Z_1 beschrieben.

$$Z_1: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 12\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq 2^{-2} \\ i: 2^{i-4} & < RUL \leq 2^{i-3} \\ 12: 2^8 & < RUL \end{cases} \quad (C.1)$$

- Die Klassengrenzen werden manuell zusammengestellt.

$$Z_2: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 10\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq 0.125 \\ 2: 0.125 & < RUL \leq 1 \\ 3: 1 & < RUL \leq 3 \\ 4: 3 & < RUL \leq 5 \\ 5: 5 & < RUL \leq 14 \\ 6: 14 & < RUL \leq 28 \\ 7: 28 & < RUL \leq 61 \\ 8: 61 & < RUL \leq 122 \\ 9: 122 & < RUL \leq 244 \\ 10: 244 & < RUL \end{cases} \quad (C.2)$$

- Klassen unterhalb von zwei Tagen RUL werden zusammengefasst sowie Restnutzungsdauern von mehr als 64 Tagen.

$$Z_3: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 6\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq 2^1 \\ i: 2^{i-1} & < RUL \leq 2^i \\ 6: 2^6 & < RUL \end{cases} \quad (C.3)$$

- Sehr wenige Klassen die nur zwei Stunden sowie drei, sechs und 60 Tage unterscheiden.

$$Z_4: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 5\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq \frac{2}{24} \\ 2: \frac{2}{24} & < RUL \leq 3 \\ 3: 3 & < RUL \leq 6 \\ 4: 6 & < RUL \leq 60 \\ 5: 60 & < RUL \end{cases} \quad (C.4)$$

- Äquidistante Klassen im Abstand von fünf Tagen.

$$Z_5: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 18\}$$

$$\text{und } z(RUL) = i: (i-1) \cdot 5 \leq RUL \leq i \cdot 5 \quad (C.5)$$

- Bei kleiner RUL wird eine geometrische Folge der Zweipotenzen genutzt und ab acht Tagen äquidistante Schritte.

$$Z_6: RUL \rightarrow I \quad \text{mit } i \in I = \{1, 2, \dots, 18\}$$

$$\text{und } z(RUL) = \begin{cases} 1: 0 & \leq RUL \leq 2^{-3} \\ i: 2^{i-5} & < RUL \leq 2^{i-4} \\ 8: 8 & < RUL \leq 16 \\ 9: 16 & < RUL \leq 24 \\ 10: 24 & < RUL \leq 32 \\ 11: 32 & < RUL \leq 40 \\ 12: 40 & < RUL \leq 48 \\ 13: 48 & < RUL \leq 56 \\ 14: 56 & < RUL \leq 64 \\ 15: 64 & < RUL \leq 72 \\ 16: 72 & < RUL \leq 80 \\ 17: 80 & < RUL \leq 88 \\ 18: 88 & < RUL \end{cases} \quad (C.6)$$

C.3 Ergänzende Informationen zur Störungsvorhersage unter realen Bedingungen

Im Abschnitt 6.3 wurde die Prognose von Weichenstörungen unter realen Bedingungen untersucht, um die Einsatzreife der Prädiktion festzustellen. Aufgrund der mittelmäßigen Resultate wurden Störungen auch individuell gelernt und prognostiziert. D.h. paarweise wurde jede Störung jeweils einzeln anhand einer davorliegenden Störung klassifiziert. Die Werte des MCC streuen sehr breit. Dies zeigt Abbildung C-1 ergänzend zum korrespondierenden ROC-Graph aus Abbildung 6-10.

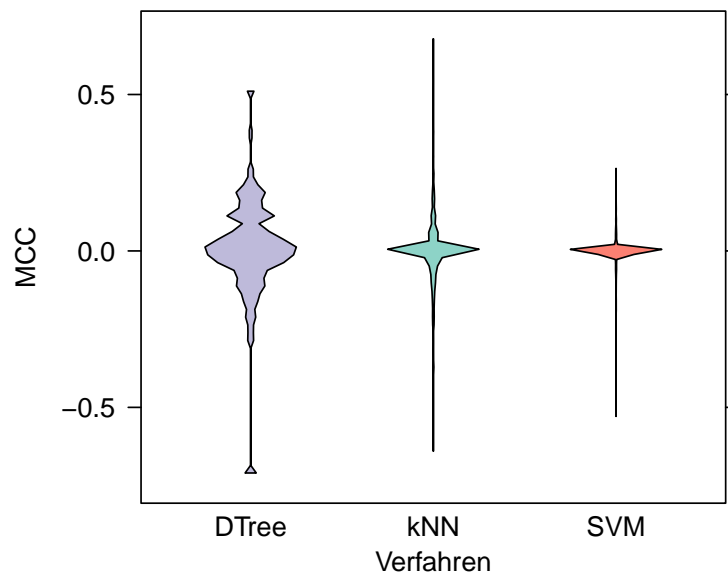


Abbildung C-1: Verteilung der MCC-Werte einer paarweisen Vorhersage unter realen Bedingungen von je einer Störung anhand jeder einzelnen zurückliegenden Störung

Aus der paarweisen Detailbetrachtung wurde auch ersichtlich, dass die Bauarteigenschaft bei einer guten Prognose eine geringere Rolle spielt als die Störungsursache selbst. Dass die richtige Bauartgruppierung das Prognoseergebniss dennoch verbessert, bestätigen die MCC-Werte. Werden bei der Prongose von Weichenstörungen die Vergangenheitsmessungen von Weichen aus demgleichen Bauartcluster herangezogen, fällt der MCC insgesamt 0.04 höher aus, als ohne die Bauartzuordnung. Die Abbildung C-2 verdeutlicht dies anhand der Ergebnisse für die Weichen. Der Effekt wird allerdings stark durch die zu geringe Menge an Störungsreferenzen überlagert.

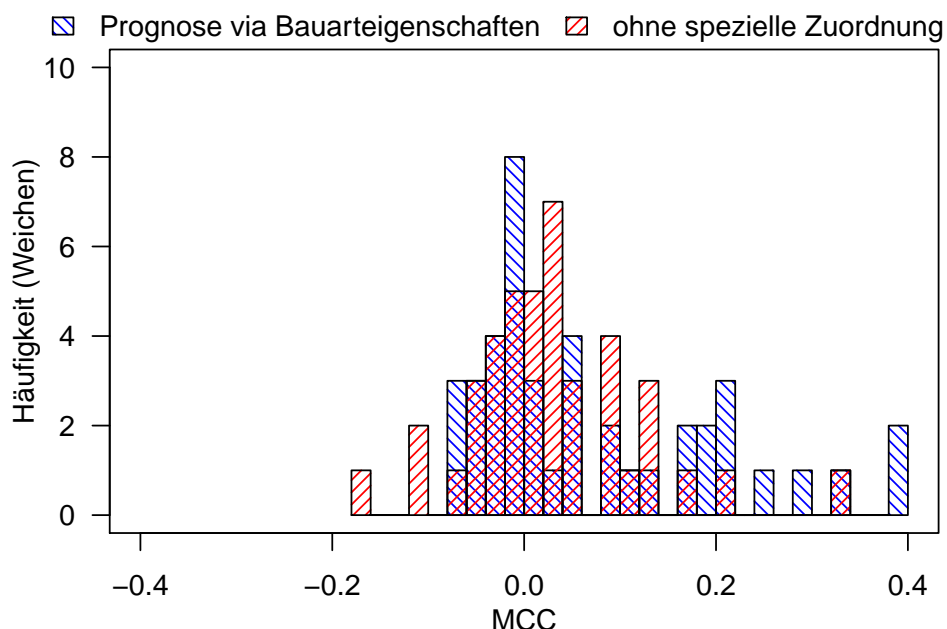


Abbildung C-2: Histogramm der MCC-Werte bei der Störungsprognose je Weiche, wenn sich die Prognose an vergangenen Störungen bei Weichen derselben Bauartgruppe orientiert und wenn nicht

Verwendete Abkürzungen

ABW Außenbogenweiche

Acc relative Genauigkeit (engl. Accuracy)

AEg Allgemeines Eisenbahngesetz

aNN künstliches Neuronales Netz (engl. artificial Neural Network)

ARMA autoregressiver, gleitender Durchschnitt (engl. Autoregressive-Moving-Average)

AT Österreich

AUC Fläche unter der ROC-Kurve (engl. Area Under the ROC Curve)

BCHB Berlin Charlottenburg

BEK Betriebserschwerungskosten

BÜ Bahnübergang

BÜSA Bahnübergangssicherungsanlage

bwgl. bewegliche

BZOO Berlin Zoologischer Garten

CBR Kosten-Nutzen-Verhältnis (engl. Cost-Benefit-Ratio)

CEN Konfusionsentropie (engl. Confusion Entropy)

CH Schweiz

cm Zentimeter

d Tage

DB AG Deutsche Bahn Aktiengesellschaft

DE Deutschland

DKW doppelte Kreuzungsweiche

DW Doppelweiche

DTree Entscheidungsbaum (engl. Decision Tree)

EBA Eisenbahn-Bundesamt

EBO	Eisenbahn-Bau- und Betriebsordnung
EIU	Eisenbahninfrastrukturunternehmen
EKW	einfache Kreuzungsweiche
ELP	Endlagenprüfer
engl.	englisch
EU	Europäische Union
EVU	Eisenbahnverkehrsunternehmen
EW	einfache Weiche
F	Kraft
FN	falsch Negative (engl. False Negatives)
FP	falsch Positive (engl. False Positives)
FR	Frankreich
GB	Großbritannien
hPa	Hektopascal
Hz	Hertz
I	elektrische Stromstärke
IBW	Innenbogenweiche
JP	Japan
KDD	Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases)
km	Kilometer
kn	Knoten
kNN	k-nächsten-Nachbarn (engl. k-Nearest-Neighbor)
LCC	Lebenszykluskosten (engl. Life Cycle Costs)
LST	Eisenbahnleit- und -sicherungstechnik
m	Meter
MCC	Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient)

mm	Millimeter
min	Minuten
Mio.	Million
MGT	Millionen Britische Tonnen (engl. Million Gross Tonnes)
MTBF	mittlere Zeit zwischen einem Ausfall (engl. Mean Time Between Failure)
MTTF	mittlere Zeit bis zu einem Ausfall (engl. Mean Time To Failure)
o. ä.	oder ähnlich
P	Leistung
PHM	Prognose und Zustandsmanagement (engl. Prognostics and Health Management)
RBF	Radial-Base-Function
ROC	Receiver Operating Characteristic
RUL	verbleibende Nutzungsdauer (engl. Remaining Useful Life)
SE	Schweden
SVM	Stützvektormaschine (engl. Support Vector Machine)
t	Tonnen
TR	Türkei
TN	richtig Negative (engl. True Negatives)
TP	richtig Positive (engl. True Positives)
U	elektrische Spannung
u.a.	unter anderem
UIC	Internationaler Eisenbahnverband (franz. Union Internationale des Chemins de Fer)
V	Volt
vgl.	vergleiche
VMin	Verspätungsminuten
wAUC	ähnlichkeitsgewichtete, makroskopische Fläche unter der ROC-Kurve (engl. weighted, macroscopic Area Under the ROC Curve)

Verwendete Formelzeichen

\hat{F}	beliebige, geschätzte, kumulierte Verteilungsfunktion, siehe Gleichung (6.8), Seite 112
\hat{f}	beliebige, geschätzte Wahrscheinlichkeitsfunktion, siehe Gleichung (6.8), Seite 112
\hat{p}	geschätzte Wahrscheinlichkeit, dass ein Sample s_l einer Klasse z_i angehört, siehe Gleichung (6.8), Seite 112
\mathbb{B}	Menge der reellen Zahlen im Intervall zwischen 0 und der maximalen Lebensdauer $maxRUL$, siehe Gleichung (4.2), Seite 65
\mathbf{C}	Konfusionsmatrix als Ergebnis einer Prognose der Samplenumenge \mathcal{S} , siehe Gleichung (4.5), Seite 65
\mathbf{K}	Kostenmatrix als Ergebnis einer Prognose der Samplenumenge \mathcal{S} , siehe Gleichung (4.6), Seite 65
\mathcal{A}	Menge von Attributen der selektierten Rohdatenbasis, siehe Gleichung (5.1), Seite 103
\mathcal{S}	Menge von selektierten Samples aller Messungen S , siehe Gleichung (4.1), Seite 65
A	Menge aller verfügbaren Attribute, siehe Gleichung (5.1), Seite 103
B	Erwarteter Nutzen einer Prognose der Samplenumenge \mathcal{S} , siehe Gleichung (4.12), Seite 68
b	Einzelner Wert der RUL, siehe Gleichung (4.2), Seite 65
B_r	Erwartete Nutzenrate einer Prognose, d.h. Nutzen je prognostiziertem Sample s , siehe Gleichung (4.13), Seite 69
B_{FP}	Kosten eines Fehlalarms, siehe Gleichung (4.12), Seite 68
B_{TP}	Nutzen einer richtig vorhergesagten Störung T , siehe Gleichung (4.12), Seite 68
C	Element der Konfusionsmatrix \mathbf{C} mit Index i und j , siehe Gleichung (4.5), Seite 65
CBR	Verhältnis der Kosten eines Fehlalarms B_{FP} zum Nutzen einer richtig vorhergesagten Störung B_{TP} , siehe Gleichung (4.16), Seite 69
F	Anzahl an Nicht-Störungen in den Mengen S oder \mathcal{S} , siehe Gleichung (4.11), Seite 68
f	beliebige Funktion, siehe Gleichung (6.3), Seite 110
F_r	Anteil an Nicht-Störungen in den Mengen S oder \mathcal{S} , siehe Gleichung (4.13), Seite 69

FP_r	Anteil der falsch Positiven (FP) an allen Nicht-Störungen, siehe Gleichung (4.10), Seite 68
G_h^v	Cluster h der Bauarteigenschaften bei gegebener Clusteranzahl v , die im hierarchischen, agglomerativen Verfahren gebildet werden, Seite 96
h	Index der Cluster der Bauarteigenschaften, Seite 96
HZ	hypothetischer Zustand, siehe Gleichung (4.4), Seite 65
I	Menge aller gebildeten Zustandsklassen, siehe Gleichung (6.2), Seite 109
i	Index der Zeile(n) bzw. der tatsächlichen Klasse in der Konfusionsmatrix \mathbf{C} oder der Kostenmatrix \mathbf{K} , siehe Gleichung (4.5), Seite 65
j	Index der Spalte(n) bzw. der hypothetischen Klasse in der Konfusionsmatrix \mathbf{C} oder der Kostenmatrix \mathbf{K} , siehe Gleichung (4.5), Seite 65
K	Element der Kostenmatrix \mathbf{K} mit Index i und j , siehe Gleichung (4.6), Seite 65
K_{Fa}	Fahrtkosten, siehe Gleichung (4.21), Seite 73
K_{FN}	Kosten einer nicht vorhergesagten Störung T , siehe Gleichung (4.20), Seite 72
K_{FP}	Kosten einer als gestört klassifizierten Nicht-Störung F , siehe Gleichung (4.21), Seite 73
K_{Pe}	Personalkosten, siehe Gleichung (4.21), Seite 73
K_{TN}	Kosten einer richtig vorhergesagten Nicht-Störung F , siehe Gleichung (4.22), Seite 73
K_{TP}	Nutzen einer richtig vorhergesagten Störung T , siehe Gleichung (4.19), Seite 72
K_{VMin}	Kostensatz je Verspätungsminute, siehe Gleichung (4.20), Seite 72
L	Menge aller Zeilen der Samplemenge \mathcal{S} , siehe Gleichung (5.2), Seite 105
l	Index der Zeile(n) bzw. Sample(s) aus der Menge \mathcal{S} , siehe Gleichung (4.1), Seite 65
m	Index der Attribute \mathcal{A} , siehe Gleichung (5.1), Seite 103
$macroAUC$	Metrik makroskopische Fläche unter der ROC-Kurve, siehe Gleichung (6.8), Seite 112
$macroFP_r$	makroskopischer Anteil der falsch Positiven (FP) an allen Nicht-Störungen, siehe Gleichung (6.7), Seite 112
$macroTP_r$	makroskopischer Anteil der richtig Positiven (TP) an allen Störungen, siehe Gleichung (6.7), Seite 112

$maxCBR$	Bestmögliches Verhältnis der Kosten eines Fehlalarms B_{FP} zum Nutzen einer richtig vorhergesagten Störung B_{TP} , siehe Gleichung (4.17), Seite 70
MCC	Metrik Matthews Korrelationskoeffizient (engl. Matthews Correlation Coefficient), siehe Gleichung (6.10), Seite 113
$minCBR$	Schlechtestes Verhältnis der Kosten eines Fehlalarms B_{FP} zum Nutzen einer richtig vorhergesagten Störung B_{TP} , siehe Gleichung (4.18), Seite 70
o	zusätzlicher Index der Konfusionsmatrix \mathbf{C} zur Berechnung des MCC, siehe Gleichung (6.10), Seite 113
p	zusätzlicher Index der Konfusionsmatrix \mathbf{C} zur Berechnung des MCC, siehe Gleichung (6.10), Seite 113
q	zusätzlicher Index der Konfusionsmatrix \mathbf{C} zur Berechnung des MCC, siehe Gleichung (6.10), Seite 113
S	Menge aller Messungen, siehe Gleichung (4.1), Seite 65
s	Einzelne Messung mit Index l , siehe Gleichung (4.1), Seite 65
T	Anzahl an Störungen in den Mengen S oder \mathcal{S} , siehe Gleichung (4.10), Seite 68
T_r	Anteil an Störungen in den Mengen S oder \mathcal{S} , siehe Gleichung (4.13), Seite 69
TP_r	Anteil der richtig Positiven (TP) an allen Störungen, siehe Gleichung (4.11), Seite 68
TZ	tatsächlicher Zustand, siehe Gleichung (4.3), Seite 65
V	Menge der möglichen Clusteranzahlen, die im hierarchischen, agglomerativen Verfahren gemäß der Bauarteigenschaften gebildet werden können, Seite 96
v	Anzahl der Cluster, die jeweil im hierarchischen, agglomerativen Verfahren gemäß der Bauarteigenschaften gebildet werden, Seite 96
$wAUC$	Metrik ähnlichkeitsgewichtete, makroskopische Fläche unter der ROC-Kurve (engl. weighted, macroscopic Area Under the ROC Curve), siehe Gleichung (6.9), Seite 113
Z	Menge der Klassen, die die RUL und damit den Zustand repräsentieren, siehe Gleichung (4.2), Seite 65
z	Einzelne Zustandsklasse aus der Menge von Zuständen Z , siehe Gleichung (4.7), Seite 67

Literaturverzeichnis

- [ACBS09] Atamuradov, V., Fatih Camci, Saim Baskan und Mehmet Sevkli: *Failure Diagnostics for Railway Point Machines Using Expert Systems*. In: *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2009*, Seiten 1–5, Piscataway (NJ USA), 2009. IEEE.
- [Ada89] Adam, Sarwat: *Optimierung der Anlageinstandhaltung: Verfügbarkeitsanforderung, Ausfallfolgekosten und Ausfallverhalten als Bestimmungsgrößen wirtschaftlich sinnvoller Instandhaltungsstrategien: Techn. Univ., Diss.–Berlin, 1989*, Band 34 der Reihe *Technological economics*. Erich Schmidt, Berlin, 1989.
- [ADA14] ADAC Online: *Monatliche Durchschnittspreise Kraftstoffe seit 2000: Kraftstoffentwicklung 2010*, 2014.
- [AH99] Adams, Niall M. und David J. Hand: *Comparing classifiers when the misallocation costs are uncertain*. *Pattern Recognition*, 32(7):1139–1147, 1999.
- [ARK13] Asada, Tomotsugu, Clive Roberts und Takafumi Koseki: *An algorithm for improved performance of railway condition monitoring equipment: Alternating-current point machine case study*. *Transportation Research Part C: Emerging Technologies*, 30(0):81–92, 2013.
- [Asa13] Asada, Tomotsugu: *Novel Condition Monitoring Techniques Applied to Improve the Dependability of Railway Point Machines*. Doktorarbeit, University of Birmingham, Birmingham, 2013.
- [Bai10] Bai, Hao: *A Generic Fault Detection and Diagnosis Approach for Pneumatic and Electric Driven Railway Assets*. Doktorarbeit, University of Birmingham, Birmingham, 2010.
- [BBC⁺00] Baldi, Pierre, Søren Brunak, Yves Chauvin, Claus A. F. Andersen und Henrik Nielsen: *Assessing the accuracy of prediction algorithms for classification: an overview*. *Bioinformatics*, 16(5):412–424, 2000.
- [BBKJ08] Böhm, Thomas, Katja Beck, Ann-Kristin Knaak und Bärbel Jäger: *Efficient maintenance strategy through System Dynamics*. In: Allan, J. J., E. Arias, C. A. Brebbia, C. J. Goodman, A. F. Rumsey, G. Sciutto und N. Tomii (Herausgeber): *Computers in railways XI*, Band XI der Reihe *WIT transactions on the built environment*, Seiten 755–764, Southampton, 2008. WIT.
- [BD10] Böhm, Thomas und Christian Doegen: *Diagnosis without Sensors - Integration of External Data for Condition Monitoring of Railway Switches*. In: Okumura, Susumu, Tadao Kawai, Peng Chen und R. BKN Rao (Herausgeber): *COMADEM 2010 - Advances in maintenance and condition diagnosis technologies towards sustainable society*, Seiten 619–622. Sunrise Publishing, Hikone (Japan), 2010.
- [Ber10] Berliner Zeitung: *Vier Tote Gleisarbeiter in zwei Tagen*. Berliner Zeitung, 21.12.2010.

- [BfV10] Verkehr, Bau und Stadtentwicklung Bundesministerium für: *Drucksache 17/3336: Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Dr. Anton Hofreiter, Winfried Hermann, Bettina Herlitzius, weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN - Zugverspätungen*, Okt 2010.
- [BG11] Böhm, Thomas und Katja Gutsche: *Diagnosis and Prediction for a Successful Management of Railway Infrastructure*. In: Singh, Maneesh, J. P. Liyanage und R. BKN Rao (Herausgeber): *COMADEM 2011*, Seiten 1413–1419, Kolkata and India, 2011. Publishing Services PL.
- [BGRS99] Beyer, Kevin, Jonathan Goldstein, Raghu Ramakrishnan und Uri Shaft: *When Is Nearest Neighbor Meaningful?* In: Beerl, Catriel (Herausgeber): *Database Theory - ICDT'99*, Band 1540 der Reihe *Lecture Notes in Computer Science*, Seiten 217–235, Berlin, 1999. Springer.
- [BHW10] Ben-Hur, Asa und Jason Weston: *A User's Guide to Support Vector Machines*. In: Carugo, Oliviero und Frank Eisenhaber (Herausgeber): *Data Mining Techniques for the Life Sciences*, Band 609 der Reihe *Methods in Molecular Biology*, Seiten 223–239. Humana Press, New York (NY USA), 2010.
- [Bie04] Biedermann, Hubert: *Leistungs- und kostenorientiertes Anlagenmanagement: Gestaltungselemente zur Kostenminimierung und Effizienzsteigerung*. Reihe Praxiswissen für Ingenieure – Instandhaltung. TÜV-Verl., Köln, 2004.
- [BJL07] Beck, Katja, Bärbel Jäger und Karsten Lemmer: *Optimisation of point life cycle costs through load-dependent maintenance*. In: Forde, M. C. (Herausgeber): *Railway Engineering 2007*, Edinburg (UK), 2007. Engineering Technics Press Edinburgh.
- [BL04] Bertsche, Bernd und Gisbert Lechner: *Zuverlässigkeit im Fahrzeug- und Maschinenbau: Ermittlung von Bauteil- und System-Zuverlässigkeiten*. VDI-Buch. Springer-Verlag Berlin Heidelberg New York, Berlin and Heidelberg, 3., überarbeitete und erweiterte Auflage Auflage, 2004.
- [Böh12a] Böhm, Thomas: *Accuracy Improvement of Condition Diagnosis of Railway Switches via External Data Integration*. In: Boller, Christian (Herausgeber): *Structural Health Monitoring 2012*, Seiten 1550–1558, Germany, 2012.
- [Böh12b] Böhm, Thomas: *Genauigkeitsverbesserung der Diagnose von Eisenbahnweichen*. EI - Eisenbahningenieur, 63.(12):56–61, 2012.
- [Böh12c] Böhm, Thomas: *Zustandsorientierte Instandhaltung bei Eisenbahnweichen*. In: Lemmer, Karsten (Herausgeber): *Next Generation Railway System*, Band 18 der Reihe *Berichte aus dem DLR-Institut für Verkehrssystemtechnik*, Seiten 51–62. DLR-Institut für Verkehrssystemtechnik, Braunschweig, 2012.
- [Böh13a] Böhm, Thomas: *Data Mining - Was nützen Statistik und künstliche Intelligenz in der Instandhaltung?*, 03.06.2013.
- [Böh13b] Böhm, Thomas: *How Precise Has Fault Detection to Be? Answers from an Economical Point of View*. In: Rao, R. BKN (Herausgeber): *COMADEM 2013*, Seiten

- 460–466, Helsinki (Finnland), 2013. KP-Media Oy.
- [Böh13c] Böhm, Thomas: *Fachworkshop Weiche: mit Instandhaltungs- und Entstörungspersonal der DB Netz AG*, 22.04.2013.
- [Böh13d] Böhm, Thomas: *Datengetriebene Zustandsdiagnose für die Instandhaltung*, 23.10.2013.
- [Bol07] Bollrath, Bernhard: *Bedienungsanleitung Sidis W: Software Sidis W Programmversion 5.28*, 2007.
- [BR02] BSL Management Consultants und R+R Burger und Partner AG: *InfraCost: The Cost of Railway Infrastructure: Final Report*, June 2002.
- [Bra97] Bradley, Andrew P.: *The use of the area under the ROC curve in the evaluation of machine learning algorithms*. Pattern Recognition, 30:1145–1159, 1997.
- [Bra07] Bramer, Max: *Principles of Data Mining*. Undergraduate topics in computer science. Springer-Verlag London Limited, London, 2007.
- [BS06] Bollrath, B. und U. Sattler: *Weichendiagnosesystem Sidis W – Betriebliche Erfahrungen*. ETR - Eisenbahntechnische Rundschau, 55(10):734–735, 2006.
- [BS10] Böhm, Thomas und Benedikt Scheier: *Railonomics - für eine bedarfsgerechte Infrastruktur: Softwaregestützte Integration von Betriebssimulation und Wirtschaftlichkeitsanalyse*. EI - Eisenbahningenieur, 61.(01):32–36, 2010.
- [Bun93] Bundesministerium der Justiz: *Allgemeines Eisenbahngesetz: AEG*, 1993.
- [Bun08] Bundesministerium der Finanzen: *Umsatzsteuer-Umrechnungskurse; Gesamtübersicht für das Jahr 2007*, Januar 2008.
- [Bun11] Bundesministerium der Finanzen: *Umsatzsteuer-Umrechnungskurse; Gesamtübersicht für das Jahr 2010*, Februar 2011.
- [Bun12] Bundesministerium der Justiz: *Eisenbahn-Bau- und Betriebsordnung: EBO*, 2012.
- [Bur98] Burges, Christopher J. C.: *A Tutorial on Support Vector Machines for Pattern Recognition*. Data Mining and Knowledge Discovery, 2:121–167, 1998.
- [Bus09] Busemann, Andreas: *Die beste Technik für das beste Netz*. Deine Bahn, 37(3):2–7, 2009.
- [CK09] Chattopadhyay, G. und S. Kumar: *Parameter Estimation for Rail Degradation Model*. International Journal of Performability Engineering, Vol. 5(2):119–130, 2009.
- [CSAA08] Chamroukhi, F., A. Samé, P. Aknin und M. Antoni: *Switch mechanism diagnosis using a pattern recognition approach*. In: *The 4th IET International Conference on Railway Condition Monitoring*, Seiten 1–4, Derby (UK), 2008.
- [CSAA11] Chamroukhi, F., A. Samé, P. Aknin und M. Antoni: *A dynamic probabilistic modeling of railway switches operating states*. In: UIC (Herausgeber): *Proceedings of*

- the 9th World Congress on Railway Research 2011*, Seiten 1–6, 2011.
- [DB 99] DB Netz AG: *Richtlinie 836: Erdbauwerke und sonstige geotechnische Bauwerke planen, bauen und instand halten*, 1999.
- [DB 02] DB Netz AG: *Richtlinie 892: LST - Anlagen montieren und instandhalten*, 2002.
- [DB 09] DB Netz AG: *Richtlinie 824: Oberbauarbeiten durchführen*, 2009.
- [DB 10] DB Netz AG: *Richtlinie 821: Oberbau inspizieren*, 2010.
- [DB 11] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2010: Internet Version*, April 2011.
- [DB 12a] DB AG: *Wettbewerbsbericht 2012*, April 2012.
- [DB 12b] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2011: Internet Version*, April 2012.
- [DB 13] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2012: Internet Version*, April 2013.
- [DB 14a] DB Netz AG: *Geschäftsbericht 2013*, April 2014.
- [DB 14b] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2013: Internetversion*, April 2014.
- [DB 16] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2015: Internetversion*, April 2016.
- [DH06] Drummond, Chris und Robert C. Holte: *Cost curves: An improved method for visualizing classifier performance*. *Machine Learning*, 65:95–130, 2006.
- [DHS12] Duda, Richard O., Peter E. Hart und David G. Stork: *Pattern Classification*. Wiley-Interscience, s.l., 2. Aufl. Auflage, 2012.
- [DIN12] DIN Deutsches Institut für Normung: *DIN EN 13306:2010-12 Instandhaltung - Begriffe der Instandhaltung; Dreisprachige Fassung*, 2010-12.
- [DIN06] DIN Deutsches Institut für Normung: *DIN 31051:2003-06 Grundlagen der Instandhaltung*, 2003-06.
- [DIN03] DIN Deutsches Institut für Normung: *DIN EN 60300-3-3:2005-03 Zuverlässigkeitsmanagement - Teil 3-3: Anwendungsleitfaden - Lebenszykluskosten; Deutsche Fassung*, 2005-03.
- [DK08] Durrant, Robert J. und Ata Kabán: *When is 'nearest neighbour' meaningful: A converse theorem and implications*. *Journal of Complexity*, 25(4):385–397, 2008.
- [Dom99] Domingos, Pedro: *MetaCost: A General Method for Making Classifiers Cost-Sensitive*. In: *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '99, Seiten 155–164, New York

- (NY USA), 1999. ACM.
- [ECG⁺11] Eker, Omer Faruk, Fatih Camci, Adem Guclu, Halis Yilboga, Mehmet Sevkli und Saim Baskan: *A simple state-based prognostic model for railway turnout systems*. IEEE Transactions on Industrial Electronics, 58(5):1718–1726, 2011.
 - [ECK10] Eker, Omer Faruk, Fatih Camci und Uday Kumar: *Failure Diagnostics on Railway Turnout Systems Using Support Vector Machines*. In: Kumar, Uday (Herausgeber): *eMaintenance 2010*, Seiten 248–251, Lulea (Sweden), 2010. Univ.
 - [Ega75] Egan, James P.: *Signal Detection Theory and ROC Analysis*. Series in Cognition and Perception. Academic Press, New York (NY USA), 1975.
 - [Faw03] Fawcett, Tom: *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*, 2003.
 - [Fen07] Fendrich, Lothar (Herausgeber): *Handbuch Eisenbahninfrastruktur*, Band 10 der Reihe *Handbuch Eisenbahninfrastruktur*. Springer and Springer-Verlag Berlin Heidelberg, Berlin and Heidelberg, 2007.
 - [FHOM09] Ferri, C., J. Hernández-Orallo und R. Modroiu: *An experimental comparison of performance measures for classification*. Pattern Recognition Letters, 30(1):27–38, 2009.
 - [FHOS03] Ferri, C., J. Hernández-Orallo und M. A. Salido: *Volume under the ROC surface for multi-class problems*. In: *Proc. of 14th European Conference on Machine Learning*, Seiten 108–120, 2003.
 - [Fie99] Fiedler, Joachim: *Bahnwesen: Planung, Bau und Betrieb von Eisenbahnen, S-, U-, Stadt- und Straßenbahnen*. Werner-Ingenieur-Texte. Werner, Düsseldorf, 4., neubearb. und erw. Aufl. Auflage, 1999.
 - [Fis84] Fischer, Klaus: *Zuverlässigkeits- und Instandhaltungstheorie*, Band 1 der Reihe *Verkehrssicherungstechnik*. Transpress Verlag für Verkehrswesen, Berlin, 1. Aufl. Auflage, 1984.
 - [FNT11] Fenner, Wolfgang, Peter Naumann und Jochen Trinckauf: *Bahnsicherungstechnik*. John Wiley & Sons, Hoboken, 2011.
 - [FPSS96a] Fayyad, Usama M., Gregory Piatetsky-Shapiro und Padhraic Smyth: *From Data Mining to Knowledge Discovery: An Overview*. In: Fayyad, Usama M., Gregory Piatetsky-Shapiro, Padhraic Smyth und Ramasamy Uthurusamy (Herausgeber): *Advances in knowledge discovery and data mining*, Seiten 1–34. AAAI Press [u.a.], Menlo Park (CA USA), 1996.
 - [FPSS96b] Fayyad, Usama M., Gregory Piatetsky-Shapiro und Padhraic Smyth: *From Data Mining to Knowledge Discovery in Databases*. AI Magazine, 17(3):37–54, 1996.
 - [FPSSU96] Fayyad, Usama M., Gregory Piatetsky-Shapiro, Padhraic Smyth und Ramasamy Uthurusamy (Herausgeber): *Advances in knowledge discovery and data mining*. AAAI Press [u.a.], Menlo Park (CA USA), 1996.

- [Fri04] Fritz, Peter: *Effiziente Diagnose zur Verbesserung der Weichen- und Streckenverfügbarkeit*. Signal + Draht, 96(11):24–27, 2004.
- [Fri09] Friedrich, Nicole: *Instandhaltung bei der DB Netz AG*. Deine Bahn, 37(Januar):20–23, 2009.
- [FS12] Fiedler, Joachim und Wolfgang Scherz: *Bahnwesen: Planung, Bau und Betrieb von Eisenbahnen, S-, U-, Stadt- und Straßenbahnen*. Werner Verlag, Köln, 6., neu bearb. und erw. Auflage, 2012.
- [GB11] Gutsche, Katja und Thomas Böhm: *e-Maintenance of Railway Assets Based on a Reliable Condition Prediction*. International Journal of Performability Engineering, Vol. 7(6):573–582, 2011.
- [GG10] Gesquière, Mickaël und Dominique Gardin: *Asset Management Line comparison*, 27.10.2010.
- [GKVJ13] Galar, Diego, Uday Kumar, Roberto Villarejo und Carl-Anders Johansson: *Hybrid prognosis for railway health assessment: an information fusion approach for PHM deployment*. In: Zio, Enrico und Piero Baraldi (Herausgeber): *2013 Prognostic and System Health Management Conference*, Band 33 der Reihe *Chemical engineering transactions*, Seiten 769–774, Milan (Italy), 2013. AIDIC Servizi S.r.l.
- [GMLTR08] García Márquez, Fausto Pedro, Richard W. Lewis, Andrew M. Tobias und Clive Roberts: *Life cycle costs for railway condition monitoring*. Transportation Research Part E: Logistics and Transportation Review, 44(6):1175–1187, 2008.
- [GMPR10] García Márquez, Fausto Pedro, Diego J. Pedregal und Clive Roberts: *Time series methods applied to failure prediction and detection*. Reliability Engineering & System Safety, 95(6):698–703, 2010.
- [GMRT10] García Márquez, Fausto Pedro, Clive Roberts und Andrew M. Tobias: *Railway Point Mechanisms: Condition Monitoring and Fault Detection*. Proceedings of The Institution of Mechanical Engineers Part F: Journal of Rail and Rapid Transit, 224(1):35–44, 2010.
- [GMWR07] García Márquez, Fausto Pedro, Paul Weston und Clive Roberts: *Failure analysis and diagnostics for railway trackside equipment*. Engineering Failure Analysis, 14(8):1411–1426, 2007.
- [Goo14] Google: *Google Maps: Routen Berlin Bellevue, Germany - Bahnhof Zoologischer Garten, Berlin bzw. Berlin-Charlottenburg, Charlottenburg*, 2014.
- [Gow71] Gower, C. J.: *A general coefficient of similarity and some of its properties*. Biometrics, 27(4):857–871, 1971.
- [GR06] Goodall, R. M. und Clive Roberts: *Concepts and techniques for railway condition monitoring*. IET Seminar Digests, 2006(11575):90–95, 2006.
- [Gra10] Gramer, Matthias: *Die Instandhaltung im Rahmen der 3-i Strategie bei der DB Netz AG*. ETR - Eisenbahntechnische Rundschau, 59(04):166–169, 2010.

- [Gün10] Günther, Oliver: *Richtlinie 420.9001: Bahnbetrieb - Kodierung der Zusatzverspätungen*, 2010.
- [Gut08] Gutsche, Katja: *Erfahrungen mit dem Einsatz von Weichendiagnosesystemen: Interview mit dem Referent Regionale Instandhaltung Entstörung DB Netz AG*, 24.09.2008.
- [Gut10] Gutsche, Katja: *Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik*. In: Lemmer, Karsten (Herausgeber): *Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik*, Band 9 der Reihe *Berichte aus dem DLR-Institut für Verkehrssystemtechnik*. DLR-Institut für Verkehrssystemtechnik, Braunschweig, 2010.
- [GYE⁺10] Guclu, Adem, Halis Yilboga, Omer Faruk Eker, Fatih Camci und Ian Jennions: *Prognostics with Autoregressive Moving Average for Railway Turnouts*. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, Seiten 1–6, 2010.
- [HAK00] Hinneburg, Alexander, Charu C. Aggarwal und Daniel A. Keim: *What Is the Nearest Neighbor in High Dimensional Spaces?* In: Abbadi, Amr el, Michael L. Brodie, Sharma Chakravarthy, Umeshwar Dayal, Nabil Kamel, Gunter Schlageter und Kyu-Young Wang (Herausgeber): *VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt*, Seiten 506–515, Orlando (Fla USA), 2000. Morgan Kaufmann.
- [Han97] Hand, David J.: *Construction and assessment of classification rules*. Wiley series in probability and mathematical statistics. Wiley, Chichester, 1997.
- [Han09] Hand, David J.: *Measuring classifier performance: a coherent alternative to the area under the ROC curve*. *Machine Learning*, 77(1):103–123, 2009.
- [HBV01] Halkidi, Maria, Yannis Batistakis und Michalis Vazirgiannis: *On Clustering Validation Techniques*. *Journal of Intelligent Information Systems*, 17(2-3):107–145, 2001.
- [HCL03] Hsu, Chih-Wei, Chih-Chung Chang und Chih-Jen Lin: *A Practical Guide to Support Vector Classification*, 2003.
- [Hem06] Hempe, Thomas: *Ein LCC-basiertes Verfahren zur Evaluierung von Schleifstrategien für Schienenbahnen*, Band 67 der Reihe *Wissenschaftliche Arbeiten für den Schienenverkehr*. Eurailpress Tetzlaff-Hestra, Hamburg, 2006.
- [Hes02] Hess, A.: *Prognostics, from the need to reality - from the fleet users and PHM system designer/developers perspectives*. In: Profet, Robert A. (Herausgeber): *2002 IEEE Aerospace Conference proceedings*, Band 6, Seiten 6–2791–6–2797 vol.6, Piscataway (NJ USA), 2002. IEEE Operations Center.
- [HG96] Hilden, Jørgen und Paul Glasziou: *Regret Graphs, Diagnostic Uncertainty and Youden's Index*. *Statistics in Medicine*, 15(10):969–986, 1996.
- [HH10] Holzfeind, Jochen und Robert Hummitzsch: *Zur Prognostizierbarkeit des Quali-*

- tätsverhaltens von Gleisen*. EI - Eisenbahningenieur, 61.(8):32–40, 2010.
- [HM82] Hanley, James A. und Barbara J. McNeil: *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. Radiology, (143(1)):29–36, 1982.
- [HMS01] Hand, David J., Heikki Mannila und Padhraic Smyth: *Principles of data mining*. A Bradford book. MIT Press, Cambridge (Mass. USA), 2001.
- [Hol12] Holst, Söhnke: *Lösung zur Instandhaltung signaltechnischer Anlagen*. Deine Bahn, 40(2):19–23, 2012.
- [Hom10] Homeyer, Dietmar: *Wissenswertes über Weichen: Teil 1*. Deine Bahn, 38(7):46–52, 2010.
- [HT01] Hand, David J. und Robert J. Till: *A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems*. Machine Learning, 45(2):171–186, 2001.
- [ISO03] ISO International Organization for Standardization: *ISO 17359:2003-11 Condition monitoring and diagnostics of machines - General guidelines*, 2003.
- [Jan09] Janicki, Jürgen: *Weichen und Kreuzungen*. Deine Bahn, 37(Januar):51–53, 2009.
- [JF10] Jurman, Giuseppe und Cesare Furlanello: *A unifying view for performance measures in multi-class prediction*. ArXiv e-prints, (Aug):1–5, 2010.
- [JRFZ12] Jurman, Giuseppe, Samantha Riccadonna, Cesare Furlanello und Giuseppe Biondi Zoccai: *A Comparison of MCC and CEN Error Measures in Multi-Class Prediction*. PLoS ONE, 7(8):e41882, 2012.
- [JZ02] Jovanovic, Stanislav und Willem-Jan Zwanenburg: *Switches and Crossings Management System: EcoSwitch: Feasibility Study*, 2002.
- [KBR06] Kalgren, P. W., C. S. Byington und M. J. Roemer: *Defining PHM, A Lexical Evolution of Maintenance and Logistics*. In: IEEE Instrumentation and Measurement Society, IEEE Aerospace and Electronic Systems Society und Institute of Electrical and Electronics Engineers (Herausgeber): *2006 IEEE Autotestcon Proceedings*, Band xxvi, Seiten 353–358, Piscataway (NJ USA), 2006. IEEE Operations Center.
- [Kne02] Knewitz, Rainer: *2. Internationaler S+D-Kongress: Verfügbarkeit – Diagnose – Instandhaltung*. Signal + Draht, 94(12):26–37, 2002.
- [Kös08] Köster, Frank: *Datenbasierte Kompetenz- und Verhaltensanalyse: Anwendungsbeispiele im selbstorganisierten eLearning: Univ., Habil.-Schr., 2007–Oldenburg, 2007*, Band 3 der Reihe *Oldenburg computer science series*. OIWR Oldenburger Verl. für Wirtschaft Informatik und Recht, Edeweicht, 2008.
- [KR11] Körkemeier, Heiner und Rolf Robbe: *Weichendiagnose Switchguard Sidis W compact*. Signal + Draht, 103(4):6–10, 2011.
- [KSBM01] Keerthi, S. S., S. K. Shevade, C. Bhattacharyya und K. R. K. Murthy: *Improve-*

- ments to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13(3):637–649, 2001.
- [Lac14] Lackhove, Christoph: *Optimierung der Migration von ETCS*. In: Lemmer, Karsten (Herausgeber): *Optimierung der Migration von ETCS*, Band 25 der Reihe *Berichte aus dem DLR-Institut für Verkehrssystemtechnik*. DLR-Institut für Verkehrssystemtechnik, Braunschweig, 2014.
- [LB11] Linoff, Gordon S. und Michael J. A. Berry: *Data mining techniques: For marketing, sales, and customer relationship management*. Timely. practical. reliable. Wiley, Indianapolis and Ind., 3. ed. Auflage, 2011.
- [LBS⁺11] Lackhove, Christoph, Florian Brinkmann, Benedikt Scheier, Ikedichi Mbakwe und Thomas Böhm: *Advancing life-cycle-management for railway signalling and control systems*. In: Finkbeiner, Matthias (Herausgeber): *Towards Life Cycle Sustainability Management*, Dordrecht, 2011. Springer Science+Business Media B.V.
- [LD07] Landgrebe, Thomas C. W. und Robert P. W. Duin: *Approximating the multiclass ROC by pairwise analysis*. *Pattern Recognition Letters*, 28(13):1747–1758, 2007.
- [Lee10] Lee, Jay: *Design of SelfMaintenance and Engineering Immune Systems for Smarter Machines and Manufacturing Systems*. In: Okumura, Susumu, Tadao Kawai, Peng Chen und R. BKN Rao (Herausgeber): *COMADEM 2010 - Advances in maintenance and condition diagnosis technologies towards sustainable society*, Seiten 1–13. Sunrise Publishing, Hikone (Japan), 2010.
- [LGY11] Leão, B. P., J. P. P. Gomes und T. Yoneyama: *Improvements on the offline performance evaluation of fault prognostics methods*. In: *IEEE Aerospace Conference, 2011*, Seiten 1–6, Piscataway (NJ USA), 2011. IEEE.
- [LRLT09] Lindtner, Frank, Ulrich Rentschler, Xiaoxi Luo und Simon Treiber: *LUPSE Online Kurs Kapitel 3: Zugsicherung: Fahrwegsicherung*, 2009.
- [LYRF08] Leão, B. P., T. Yoneyama, G. C. Rocha und K. T. Fitzgibbon: *Prognostics performance metrics and their relation to requirements, design, verification and cost-benefit*. In: *International Conference on Prognostics and Health Management, 2008*, Seiten 1–8, Piscataway (NJ USA), 2008. IEEE.
- [Mat75] Matthews, B. W.: *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. *Biochimica et biophysica acta*, 405(2):442–451, 1975.
- [Mil85] Milligan, Glenn: *An algorithm for generating artificial test clusters*. *Psychometrika*, 50(1):123–127, 1985.
- [ML05] Min, Jae H. und Young-Chan Lee: *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters*. *Expert Systems with Applications*, 28(4):603–614, 2005.
- [MR05] Maimon, Oded und Lior Rokach: *Decomposition methodology for knowledge discovery and data mining: Theory and applications*, Band 61 der Reihe *Series in machine perception and artificial intelligence*. World Scientific, New Jersey, 2005.

- [MR10] Maimon, Oded und Lior Rokach (Herausgeber): *Data Mining and Knowledge Discovery Handbook*. Springer Science+Business Media LLC, Boston (MA USA), 2. ed. Auflage, 2010.
- [NP04] Naumann, P. und J. Pachl (Herausgeber): *Leit- und Sicherungstechnik im Bahnbetrieb*. Tetzlaff, Hamburg, 2 Auflage, 2004.
- [Pac09] Pachl, Jörn: *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*. Vieweg+Teubner (GWV), s.l., 5. Aufl. Auflage, 2009.
- [Pet05] Petersohn, Helge: *Data Mining: Verfahren, Prozesse, Anwendungsarchitektur: Univ., Habil–Leipzig, 2004*. Oldenbourg, München and Wien, 2005.
- [PFK98] Provost, F., T. Fawcett und R. Kohavi: *The Case Against Accuracy Estimation for Comparing Inductive Algorithms*. In: *Proceedings of the Fifteenth International Conference on Machine Learning*, Seiten 445–453, 1998.
- [PGMR09] Pedregal, Diego J., Fausto Pedro García Márquez und Clive Roberts: *An algorithmic approach for maintenance management based on advanced state space systems and harmonic regressions*. Annals of Operations Research, 166(1):109–124, 2009.
- [PGMS04] Pedregal, Diego J., Fausto Pedro García Márquez und Felix Schmidt: *RCM² predictive maintenance of railway systems based on unobserved components models*. Reliability Engineering & System Safety, 83(1):103–110, 2004.
- [Pla99] Platt, John C.: *Fast Training of Support Vector Machines Using Sequential Minimal Optimization: Advances in Kernel Methods*. In: Schölkopf, Bernhard, Christopher J. C. Burges und Alexander J. Smola (Herausgeber): *Advances in Kernel Methods*, Seiten 185–208. MIT Press, Cambridge (Mass. USA), 1999.
- [Pöl10] Pölinger, Alfred: *Weichen mit integrierten Stellsystemen*. Signal + Draht, 102(1+2):25–28, 2010.
- [Qui93] Quinlan, John R.: *C4.5: programs for machine learning*. The Morgan Kaufmann series in machine learning. Morgan Kaufmann Publishers Inc., San Mateo (CA USA), 5. [pr.] Auflage, 1993.
- [Qui11] Quiroga, Lisandro M.: *Ganzheitliche Optimierung des Instandhaltungsprozesses der Gleisgeometrie*. Doktorarbeit, Technische Universität Braunschweig, Braunschweig, 2011.
- [QWS09] Quiroga, Lisandro M., Stefan Wegele und Eckehard Schnieder: *Benefit of railway infrastructure diagnosis systems on its availability*. IFAC Proceedings Volumes, 42(5):146–150, 2009.
- [RA13] Rama, Dovile und John D. Andrews: *A reliability analysis of railway switches*. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 227(4):344–363, 2013.
- [Rau10] Rausch, Wolfgang: *Diagnosesysteme für Weichen als Grundlage für eine optimierte Instandhaltungsplanung*. In: VDEI (Herausgeber): *Symposium zum Thema: „Mo-*

- derne Instandhaltungsverfahren für Weichen – Qualitätsansprüche – Wirtschaftlichkeit“. Brandenburg-Kirchmöser, 2010.
- [RB93] Riedmiller, M. und H. Braun: *A direct adaptive method for faster backpropagation learning: theRPROP algorithm*. In: *Proceedings of the IEEE International Conference on Neural Networks*, Seiten 586–591, Piscataway (NJ USA), 1993.
- [RB03] Redeker, Frank Robert und Dionijs Emil Brugman: *Ferndiagnose von Eisenbahnobjekten*. *EI - Eisenbahningenieur*, 54(03):16–20, 2003.
- [RM10] Rokach, Lior und Oded Maimon: *Classification Trees*. In: Maimon, Oded und Lior Rokach (Herausgeber): *Data Mining and Knowledge Discovery Handbook*, Seiten 149–174. Springer Science+Business Media LLC, Boston (MA USA), 2010.
- [SAM96] Shafer, John C., Rakesh Agrawal und Manish Mehta: *SPRINT: A Scalable Parallel Classifier for Data Mining*. In: Vijayaraman, T. M. (Herausgeber): *Very large data bases, VLDB '96*, Seiten 544–555, San Francisco (CA USA), 1996. Morgan Kaufmann.
- [SAR10] Sebastiani, Paola, Maria M. Abad und Marco F. Ramoni: *Bayesian Networks*. In: Maimon, Oded und Lior Rokach (Herausgeber): *Data Mining and Knowledge Discovery Handbook*, Seiten 175–208. Springer Science+Business Media LLC, Boston (MA USA), 2010.
- [SB02] Stoll, Horst und Bernhard Bollrath: *Weichendiagnosesystem SIDIS W*. *Signal + Draht*, 94(4):26–29, 2002.
- [SB14] Schenkendorf, René und Thomas Böhm: *Aspekte einer datengetriebenen zustandsabhängigen Instandhaltung: (Teil 1) Informative Merkmalsextraktion als Basis einer zuverlässigen Zustandsdiagnose*. *EI - Eisenbahningenieur*, 65(11):14–18, 2014.
- [SCB⁺08] Saxena, Abhinav, J. Celaya, E. Balaban, Kai Goebel, B. Saha, S. Saha und M. Schwabacher: *Metrics for Evaluating Performance of Prognostic Techniques*. In: *International Conference on Prognostics and Health Management, 2008. PHM 2008.*, Seiten 1–17, 2008.
- [Sch05] Schild, Ulrich: *Lebenszyklusrechnung und lebenszyklusbezogenes Zielkostenmanagement: Stellung im internen Rechnungswesen, Rechnungsausgestaltung und modellgestützte Optimierung der intertemporalen Kostenstruktur: Univ., Diss.–Göttingen, 2004*. Gabler-Edition Wissenschaft. Dt. Univ.-Verl., Wiesbaden, 1. Aufl. Auflage, 2005.
- [Sch09] Schönemann, René: *Methoden und Kriterien zur Bewertung von Eisenbahninfrastruktur*. Doktorarbeit, Technische Universität Dresden, Dresden, 2009.
- [SCS⁺10] Saxena, Abhinav, J. Celaya, B. Saha, S. Saha und Kai Goebel: *Evaluating prognostics performance for algorithms incorporating uncertainty estimates*. In: *IEEE Aerospace Conference, 2010*, Seiten 1–11, Piscataway (NJ USA), 2010. IEEE.
- [Sei08] Seidl, Kurt: *Neue Diagnosesysteme verringern die Instandhaltungskosten*. *Signal + Draht*, 100(11):26–30, 2008.

- [Shm10] Shmilovici, Armin: *Support Vector Machines*. In: Maimon, Oded und Lior Rokach (Herausgeber): *Data Mining and Knowledge Discovery Handbook*, Seiten 231–247. Springer Science+Business Media LLC, Boston (MA USA), 2010.
- [Sie12] Siemens Schweiz AG: *Switchgard CKA: Weichen für die Zukunft*, 2012.
- [Sil09] Silmon, Joseph A.: *Operational Industrial Fault Detection and Diagnosis: Railway Actuator Case Studies*. Doktorarbeit, University of Birmingham, Birmingham, 2009.
- [SKW08] Sheppard, John W., Mark A. Kaufman und Timothy J. Wilmering: *IEEE standards for prognostics and health management*. In: *IEEE Autotestcon, 2008*, Seiten 97–103, Piscataway (NJ USA), 2008. IEEE.
- [SL03] Schilling, Rosemarie und Lars Lücking: *Senkung der Lebenszykluskosten: Standardisierung von Instandhaltungs- und Umbaumaßnahmen am Beispiel der Weichen*. EI - Eisenbahningenieur, 54(5):58–72, 2003.
- [SL08] Sostaric, Georg und Christian Langschwert: *Doppelte Ausnutzung modularer Weichenantriebssysteme*. Signal + Draht, 100(5):11–15, 2008.
- [SL09] Sokolova, Marina und Guy Lapalme: *A systematic analysis of performance measures for classification tasks*. Information Processing and Management, 45(4):427–437, 2009.
- [SLB14] Schenkendorf, René, Christian Linder und Thomas Böhm: *Potenziale, Techniken und Algorithmen für die Zustandsdiagnose und -prognose bei LST-Elementen*. In: Verband Deutscher Eisenbahn-Ingenieure e.V. - VDEI (Herausgeber): *EIK Eisenbahningenieurkalender 2015*, Seiten 159–172. DVV Media Group GmbH | Eurailpress, Hamburg, 2014.
- [Stu96] Sturm, Adolf: *Zustandswissen für Betriebsführung und Instandhaltung*, Band Band 10 der Reihe *Fachbuchreihe "Kraftwerkstechnik"*. VGB-Kraftwerkstechnik GmbH, Essen, 1. Ausg. Auflage, 1996.
- [Swe88] Swets, John A.: *Measuring the accuracy of diagnostic systems*. Science, 240:1285–1293, 1988.
- [Thi10] Thiel, Hans-Christoph: *Erhaltungswirtschaftliche Aspekte der Planung und des Einsatzes von Weichen*. In: VDEI (Herausgeber): *Symposium zum Thema: „Moderne Instandhaltungsverfahren für Weichen – Qualitätsansprüche – Wirtschaftlichkeit“*. Brandenburg-Kirchmöser, 2010.
- [TK09] Theodoridis, Sergios und Konstantinos Koutroumbas: *Pattern recognition*. Elsevier/Acad. Press, Amsterdam, 4. ed. Auflage, 2009.
- [Tur95] Turney, Peter D.: *Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm*. Journal of Artificial Intelligence Research (JAIR), 2:369–409, 1995.
- [UGL08] Uckun, Serdar, Kai Goebel und Peter J.F. Lucas: *Standardizing research methods*

- for prognostics. In: *International Conference on Prognostics and Health Management, 2008*, Seiten 1–10, Piscataway (NJ USA), 2008. IEEE.
- [UIC08] UIC: *Kontinuierliches Benchmarking der Infrastrukturkosten (LICB): Synthesebericht Dezember 2008*, 2008.
- [Uni08] University of Birmingham: *INNOTRACK (FP6 EU Project): Deliverable D3.3.1 – List of key parameters for switch and crossing monitoring*, 2008.
- [Vap08] Vapnik, Vladimir N.: *The nature of statistical learning theory*. Statistics for engineering and information science. Springer, New York, 2. ed., 6. print Auflage, 2008.
- [VCD12] VCD Verkehrsclub Deutschland: *Gute Bahn – Schlechte Bahn*, 2012.
- [VDI12] VDI-Gesellschaft Produktion und Logistik: *Zustandsorientierte Instandhaltung (Maintenance condition monitoring)*, 1999–12.
- [VP08] Veit, Peter und Klaus Petri: *Betriebserschwerniskosten: ein Baustein zur Systemoptimierung*. ZEV rail Glasers Annalen, 132(5):168–175, 2008.
- [VRPRA13] Vileiniskis, Marius, Rasa Remenyte-Prescott, Dovile Rama und John Andrews: *Fault Diagnostics for Railway Point Machines*. In: Jackson, Lisa und John Andrews (Herausgeber): *Proceedings of the 20th AR2TS Advances in Risk and Reliability Technology Symposium*, Seiten 103–119, Loughborough (Leicestershire UK), 2013. Loughborough University.
- [wet09] wetter.com: *Wetterarchiv Station 10405 Berlin-Prenzlauer Berg*, 2009.
- [Wik14] Wikipedia: *Opel Astra H*, 2014.
- [Win10] Winter, Josef: *Diagnosesysteme für Weichen als Grundlage für eine optimierte Instandhaltungsplanung*. In: VDEI (Herausgeber): *Symposium zum Thema: „Moderne Instandhaltungsverfahren für Weichen – Qualitätsansprüche – Wirtschaftlichkeit“*. Brandenburg-Kirchmöser, 2010.
- [WKP10] Wheeler, Kevin R., Tolga Kurtoglu und Scott D. Poll: *A Survey of Health Management User Objectives Related to Diagnostic and Prognostic Metrics*. International Journal of Prognostics and Health Management, 1(003):1–19, 2010.
- [Wüb84] Wübbenhorst, Klaus L.: *Konzept der Lebenszykluskosten: Grundlagen, Problemstellungen und technologische Zusammenhänge: TH, Diss.–Darmstadt, 1984*, Band 7=12 der Reihe *Reihe Betriebswirtschaft*. Verlag für Fachliteratur, Darmstadt, 1984.
- [YEGC10] Yilboga, Halis, Omer Faruk Eker, Adem Guclu und Fatih Camci: *Failure prediction on railway turnouts using time delay neural networks*. In: *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSAs), 2010*, Seiten 134–137, Piscataway (NJ USA), 2010. IEEE.
- [ZAB⁺02] Zhou, F., N. Archer, J. Bowles, M. Duta, M. Henry, M. Tombs, M. Zamora,

- S. Baker und C. Burton: *Remote Condition Monitoring and Validation of Railway Points*. Computing Control Engineering Journal, 13(5):221–230, 2002.
- [Zha10] Zhang, Peter G.: *Neural Networks For Data Mining*. In: Maimon, Oded und Lior Rokach (Herausgeber): *Data Mining and Knowledge Discovery Handbook*, Seiten 419–444. Springer Science+Business Media LLC, Boston (MA USA), 2010.
- [Zwa09] Zwanenburg, Willem-Jan: *Degradation Processes of Railway Switches & Crossings: To Improve Maintenance & Renewal Planning on the Swiss Railway Network*. Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken, 2009.